2002

The representation of metric information in scenes

Luke Justin Rosielle
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

The representation of metric information in scenes

by

Luke Justin Rosielle

A dissertation submitted the to graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Psychology

Program of Study Committee:
Eric E. Cooper, Major Professor
Veronica J. Dark
William S. Robinson
Gary L. Wells
Dennis Dake

Iowa State University

Ames, Iowa

2002

UMI Number: 3073476

UMI Microform 3073476

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Luke Justin Rosielle

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT

Four experiments were conducted that investigated the role of metric information in the identification and episodic recognition of scenes. A fifth experiment examined whether scene identification showed any hemispheric advantage. For scene identification, a priming paradigm was used in which participants were required to identify scenes that were identical, that changed size, or that changed location in the visual field relative to a previously viewed scene. For episodic recognition, participants were required to indicate whether a scene had been viewed previously while ignoring any changes in size or position. The results found a reduced identification or episodic recognition advantage for previously viewed scenes that had changed their size or position in the visual field relative to scenes that were identical to one viewed previously. In addition, no evidence of laterality was found. The results challenge the notion that scenes are treated in the same way by the visual system as objects or faces.

CHAPTER 1: INTRODUCTION

One of the fundamental questions in visual cognition is how people recognize the diverse elements that make up their visual world. For example, as I am sitting at my desk typing this sentence I can easily and effortlessly recognize my computer, the books on my desk, the words I am typing, the faces of the people in the photographs scattered throughout my office, and that this place I am in is indeed my office. A substantial amount of research has been conducted to try to discover how the visual system processes and stores information in order for recognition to occur.

What is the Visual System?

What exactly constitutes the visual system? Unfortunately, there are no stringent criteria for defining what comprises the visual system (or any other brain system, for that matter) and any boundaries imposed (such as the boundary between the visual system and memory) are likely to be somewhat arbitrary and artificial. The brain seems to process visual information in the series of incremental steps and at higher levels of processing there is no clear boundary between "vision" and "memory".

To complicate matters, the visual system is composed of a number of subsystems defined loosely by the individual tasks they are designed to perform and the neural pathways that comprise them. However, these subsystems often overlap and are highly interconnected. For example, the dorsal and ventral cortical pathways in the visual system are thought to be responsible for recognition (ventral) and spatial vision (dorsal), but share several cortical areas in common and have more connections between these two pathways than within the individual pathways themselves (Milner & Goodale, 1995). In contrast Zeki (1999) proposed four visual systems based on cortical pathways: two for form perception

(one for static objects and another for moving objects), one for motion perception and one for color perception.

For the purposes of the current work, I will adopt a broad definition of the visual system. The visual system is the entire processing stream from the retina to the motor areas responsible for generating a response to visual stimuli. Because memory for objects seems to be neurologically distinct from memory for faces, memory for sounds, and other types of memories (see McCarthy & Warrington, 1990, for a review), these types of memory are more properly thought of as part of their parent system or corresponding subsystem (e.g., visual memory as part of the visual system, auditory memory as part of the auditory system, memory for faces is part of the face recognition system which itself is a part of the visual system). Visual memory (memory for the visual aspects of elements of the environment) is thus included as part of the visual system. Thus, visual recognition can be said to be accomplished by the visual system.

Visual Recognition

In a broad sense, visual recognition is the process by which visual input is matched to a stored representation in memory. Visual recognition is also the end state of the system that accomplishes this match; if visual input has been successfully matched to a stored representation in memory, then visual recognition has occurred. Thus, recognition can be thought of both as the process and a goal. It is the nature of the memory representation that allows for visual recognition that will be the subject of the experiments reported here.

More specifically, recognition has two distinct meanings in the literature. First, recognition can mean to identify something as being a member of a certain class or category. For example, recognition might mean recognizing a certain piece of visual input as a "bird"

or an "American Robin". In contrast, recognition might also mean recognizing a stimulus as a specific instance of a category that one has encountered before. For example, recognition in the second case might mean recognizing a face as belonging to someone that a person has previously met. The first type of recognition is thus context free (e.g., "that is a bird") whereas the second type of recognition is invariably tied to a context (e.g., "that is someone I met at the party last week.").

Although the word recognition is often used without reference to which type of recognition is being referred to (see for example, Christou & Bulthoff, 1999; Nakatani, Pollatsek, & Scott, 2002; Oliva & Schyns, 2000), evidence suggests that the processes that underlie each of these two types of recognition are different (Biederman & E. E. Cooper, 1992; Tulving & Schacter, 1990). Thus, when using the term recognition, it is important to specify precisely what type of recognition is being discussed.

In order to be precise, in the current work the word "identification" will be used specifically to refer to the type of recognition that involves assigning a stimulus to a certain class. Thus, identification means to recognize a stimulus as being a certain type (e.g., "that object is a bird"). "Episodic recognition" will refer specifically to the type of recognition that involves recognizing a stimulus as something specific that has been encountered previously. Episodic recognition would thus refer to recognizing an object as a specific object that one as seen previously (e.g., "I saw that bird yesterday."). "Recognition" will be used when referring to both processes.

In the empirical and theoretical literature on visual recognition, much of the focus has been directed at investigating the processes underlying the recognition of three specific classes of perceptual stimuli: individual objects (e.g., cars and chairs), people's faces (e.g.,

Tom Cruise's face and Hillary Clinton's face), and letters and words (e.g., the letter "M" or the word "bench") (see Farah, 1994, for a review). A consensus is developing that the representation used to recognize these types of stimuli varies according to the type of stimulus. For example, numerous differences have been identified between the representation used for most forms of object recognition and the representation used for face recognition, including differences in the type of visual primitives used (Bruce & Langton, 1994), the manner in which the spatial relations among the primitives are computed (E. E. Cooper & Wojan, 2000; Diamond & Carey, 1986; Tanaka & Farah, 1993), and the cortical areas responsible for object and face recognition (Davidoff, 1982; H. D. Ellis, 1983; Sergent, Ohta, & MacDonald, 1992). Thus, visual recognition may not employ a single general process for recognition, but rather may employ multiple routes to recognition depending on the type of stimulus present. The goal of much of the empirical work in visual recognition is to identify differences in the representational format for various types of stimuli and tasks.

Scene Recognition

An equally important but empirically neglected facet of a person's recognition system is the ability to recognize places in the environment (i.e., scenes). Henderson and Hollingworth (1999) defined a scene as "a semantically coherent (and often nameable) view of a real world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner" (p. 244). Thus, scenes contain both background information (e.g., walls, floor, sky) and discrete foreground objects (e.g., tables, chairs, lamps, trees) arranged in three-dimensional space. Under this definition, scenes are not simply collections of random objects or shapes, but rather are comprised of objects spatially arranged in a particular context.

The processes underlying the visual analysis of scenes have been the focus of a wide range of empirical investigations (Henderson & Hollingworth, 1999). As noted by Henderson and Hollingworth, most research on scene perception can be divided into three general categories; studies that investigate how scenes are visually scanned (e.g., Antes, 1974; Buswell, 1935; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967; Rayner & Pollatsek, 1992; Yarbus, 1967), studies that investigate how the context of a scene affects the identification of individual objects (e.g., Biederman, 1972; Biederman, Glass & Stacy, 1973; Biederman, Mezzanotte, & Rabinowitz, 1982; De Graef, De Troy, & D'Ydewalle, 1992; Palmer, 1975) , and studies that investigate the processes that underlie the detection of change in scenes (e.g., Archambault, O'Donnell, & Schyns, 1999; Henderson & Hollingworth, 1999; Rensink, 2000; Rensink, O'Regan, & Clark, 1997; Rosielle, Crabb, & E. E. Cooper, 2002).

Despite the large number of studies on scene perception, a fundamental question remains to be answered: what is the nature of the memory representation that allows scenes to be recognized? Just as other types of visual stimuli can quickly and effortlessly be identified as members of a particular category (e.g., "desk", "car", "microwave"), scenes can be quickly and effortlessly identified as belonging to a particular category as well (e.g., "kitchen", "farm", "office"). Similarly, scenes and other types of objects can also be recognized as familiar (i.e., episodic recognition); one can recognize a familiar street or living room just as one can identify a familiar coffee cup or car. Very little (if anything) is known about how the visual system accomplishes these types of recognition; in fact, Henderson and Hollingworth (1999) noted that "Future research will need to identify the

internal representations constructed from a scene and the processes by which these representation are compared to stored scene representations" (p. 266).

This lack of attention to scene recognition is somewhat surprising given that in the real world a person's visual input almost always consists of some kind of scene. While recognizing the individual elements (e.g., the objects) that comprise the scene is certainly important, recognizing the scene itself is a fundamental and critical step in a variety of visual tasks. For example, navigating through the world often requires scene recognition, as a person must recognize his or her surroundings in order to assess where he or she currently is located. In addition, knowing the type of scene one is currently viewing (e.g., "kitchen or "office") gives important information as to what specific types of objects are likely to be present and their likely locations, thus making the processing of this type of information more efficient (Antes, Penland, & Metzger, 1981; Biederman, 1981; De Graef, 1992; Friedman, 1979; Henderson, 1992; Metzger & Antes, 1983; Palmer, 1975; Rensink, 2000). Further, understanding and interpreting many kinds of visual media (e.g., some works of art, pictorial directions, movie and television images) requires one to identify the place or the type of scene the picture is depicting.

Thus, knowing the type of information people store in memory that allows scenes to be recognized provides critical information about how people accomplish a variety of important visual tasks. The research reported here is to provides some of the first steps in investigating the nature of the internal memory representation that allows scene recognition to occur.

Does Scene Identification Depend on Object Identification?

Given that scenes are more or less collections of individual objects, then why is it important to study scene identification separately from the identification of individual objects? After all, it seems plausible that scene identification may just be a matter of first recognizing a few diagnostic objects that comprise the scene and drawing a conclusion about the scene's identity based on the objects present. For example, some researchers have suggested that a person might identify an office by first recognizing a desk, a chair, and a computer and conclude they are viewing an office (Friedman, 1979). This route to scene identification is thought to be intentional, conscious, and is thought to require that spatial attention be allocated to various regions of the scene in order for identification to occur (see Henderson & Hollingworth, 2000). Critically, this route is time intensive and predicts that scene identification cannot take place before the eyes can move (saccade) to various regions of the scene, as spatial attention is often dictated by what is in foveal vision (Schyns & Oliva, 1994). If scene identification occurred in this way, one could argue that there is no reason to study scene identification separately from object identification because according to this view, all the visual system is doing during scene identification is identifying the individual objects.

Although the idea that people identify scenes by first identifying the objects is intuitive, most studies suggest that scenes are generally not initially identified in this way (see Biederman, 1988; Henderson & Hollingworth 1999; Oliva & Schyns, 2000; Schyns & Oliva, 1994, for reviews). Several lines of evidence support the view that object identification does not have to precede scene identification.

First, several authors have argued that scene identification occurs too quickly for a scene to be effectively scanned for objects diagnostic of a scene's identity to be identified and used to make an identification judgment (Biederman, 1987b; Biederman , Mezzanotte, & Rabinowitz, 1982; Henderson & Hollingworth, 2000; Oliva & Schyns, 1997; Schyns & Oliva, 1994). Scene identification can occur very quickly, taking less than 100 ms. (Potter, 1976), a duration comparable to that of individual object identification (Biederman, 1987b; Biederman et al., 1982; Potter, 1976; Schyns & Oliva, 1994), and prior to visual scanning taking place. For example, Schyns and Oliva (1994) found that a scene can be identified as a certain type (e.g., "kitchen") from a masked presentation of 45 to 135 ms, indicating that the information needed to identify a scene can be extracted very quickly and without visually scanning the scene (given that it takes approximately 250 ms for the eyes to saccade). Other research demonstrates that the identities of the objects in a scene that are not in foveal vision are often not consciously available to the viewer (Henderson & Hollingworth, 1999), but people can still identify scenes without visually scanning the scene (Loftus, Nelson, & Kallman, 1983; Oliva & Schyns, 1994; Potter, 1976; Schyns & Oliva, 1997). These findings indicate that something other than the identities of the objects in a scene is used to identify scenes.

A second line of evidence that scene identification does not depend on object identification is the finding that bottom-up object identification is not necessary for successful scene identification to occur. For example, Biederman (1988) has shown that scenes can be easily identified even when the objects that compose the scene are reduced to a set of one-part volumetric shapes (e.g., cones and cylinders). In these scenes, the objects themselves, if presented in isolation, would be unrecognizable as anything other than

volumetric shapes and yet they can be assembled to produce an identifiable scene. Similarly, Schyns and Oliva (1994) (also Oliva & Schyns, 1997) found that scenes in which high spatial-frequency information was removed, thus rendering the individual objects in the scene unrecognizable, can also be identified quickly. Similarly, entire scenes can be identified at exposure durations brief enough to render the individual objects in the scene difficult to identify (Schyns & Oliva, 1994). Further, Oliva and Schyns (Oliva & Schyns, 1997; Schyns & Oliva, 1994) found that at pre-saccade exposure durations (i.e., less than approximately 250 ms), people base their judgments of scene identity on low spatial frequency information when presented with composite scenes composed of high-spatial frequency information from one scene and low spatial frequency information from a different scene. These results have led several authors to argue that scene identification is initially based on low spatial-frequency information but not on the identity of the objects themselves (see also Antes et al., 1981; Loftus et al., 1983; Metzger & Antes, 1983 for similar views).

Thus, most evidence suggests that scene identification is not simply a matter of recognizing a set of individual objects and drawing a conclusion about the type of scene present. Rather, a representation of the scene is formed prior to object identification taking place, and it is this representation that is used for scene identification (an approach Navon, 1977, cleverly calls recognizing the "forest before the trees"; see also Schyns & Oliva, 1994).

A Common Representational Code for Scenes and Objects?

Very little is currently known, however, about the memory representation responsible for scene identification, although Biederman (1988) offered one possibility (see also

Biederman, 1987a; 1987b; Diwadkar & McNamara, 1997; Shelton & McNamara, 1997). Biederman proposed that the visual system processes an entire scene in exactly the same way as the visual system processes an individual object, storing individual scenes in memory simply as large objects. A "kitchen", for example, would be identified using the same set of processes as would a "lamp". According to this idea, the scene itself is an object and is treated as such by the visual system. That is, scenes and objects have a common representational code.

As noted by Biederman (1988), scenes in many ways are similar to individual objects, but exist on a larger spatial scale. Both scenes and objects are composed of multiple, spatially arranged parts, must be identified from different viewpoints, and share a similar time course for identification. According to Biederman, the object identification system could account for scene identification as well.

In support of Biederman's idea, recent evidence suggests that the same types of spatial relations coded in object identification are also coded in scene perception. Rosielle, Crabb, and E. E. Cooper (2002) found that people can detect categorical changes to the positions of objects in scenes (e.g., moving an object from "below" another object to "side-of" another object) faster than they can detect an equal metric change (in terms of Euclidian distance) that does not disrupt categorical relations (e.g., moving an object so that it remains "above" another object). This finding suggests that the relations among the elements in a scene are being coded categorically, a finding that mirrors numerous studies that have found a role for categorical relations in object identification (e.g., E. E. Cooper & Wojan, 2000; Hummel & Stankiewicz, 1996; Rosielle & E. E. Cooper, 2001).

Other lines of evidence, however, suggest that scenes and objects may not share a common representational code. Several studies have found evidence that the spatial relations between the parts of a single object (i.e., "within-object" relations, such as the relation between the curved handle of a mug and the cylindrical base of a mug) are coded differently than the spatial relations between multiple discrete objects (i.e., "between-object" relations, such as the relations between the cup and another object) (e.g., Baylis & Driver, 1993, Humphreys & Riddoch, 1994, 1995; Saiki and Hummel, 1998, but see Rosielle, Crabb, & E. E. Cooper, 2001; Shelton & McNamara, 1997). For example, Baylis and Driver (1993) found that people are better at judging the relative height of two vertices that belong to the same object (a judgment of within-object relations) than they are at judging the relative height of two identical vertices that belong to different objects (a judgment of between-object relations). Likewise, Saiki and Hummel (1996) found that people's same/different judgments were more affected by changes in orientation when the parts of a stimulus were disconnected compared to when the parts of a stimulus were connected. Further, Humphreys and Riddoch (1994) described a patient with unilateral spatial neglect who neglected the right side of space when computing between object relations but neglected the left side of space when computing within-object relations. They interpreted these results as evidence that the visual system codes within-object relations, which are required for object identification, separately (and perhaps differently) than between-object relations, which are required for scene identification.

Further evidence that scenes and objects are coded using different types of representations comes from neuropsychology. Epstein and Kanwisher (1998), using functional magnetic resonance imaging (fMRI), localized an area of the parahippocampal

cortex that becomes active in response to passively viewed realistic scenes, but only weakly active in response to the individual objects that comprise the scene with the background information removed. These results suggest that there may be a separate cortical area devoted to scene processing.

However, it is unclear the extent to which the results of the above mentioned studies can be applied to scene identification. Many of these experiments used abstract, two-dimensional figures with separated (non-touching) parts, such as the vertices used by Baylis and Driver (1993). It is currently unclear the extent to which figures like these constitute "scenes". For example, the "scene area" in the parahippocampal cortex only becomes active to naturalistic scenes with three-dimensional depth information, but not to collections of objects or scrambled scenes. Similarly, Zaidel and Kasher (1989) found that memory for realistic paintings did not depend upon the visual field to which the painting was projected, whereas memory for abstract paintings was affected by the visual field (people had better memory for the abstract paintings that were presented to the left visual field). These findings raised the possibility that abstract scenes (such as the disconnected line segments) may be treated differently by the visual system than more naturalistic scenes with background information. Further, the results from Epstein and Kanwisher (1998) suggested different cortical locations for scene and object perception, but leave open the question of how (and whether) the specific types of representations used for scene and object identification might differ. Finally, these studies employed varying tasks that did not test scene identification directly and thus may have produced results that do not apply to the identification of scenes.

Repetition Priming

Repetition priming is the phenomenon whereby participants will respond more quickly to a stimulus the second time the stimulus is presented than they did the first time (Bartram, 1974). In a typical repetition priming paradigm, participants make some response to a stimulus in a first block of trials (e.g., naming the stimulus, matching the stimulus to a name, or assigning the stimulus to a category). Then, in a second block of trials, participants respond to the same stimulus again. Priming is defined as the improvement in the participant's performance (usually measured as a reduction in response times and errors) from the first to the second presentation of the stimulus.

The most common explanation for the repetition priming phenomenon is that on the second presentation of the stimulus, the representation in memory of that stimulus in memory is still partially active (Reisberg, 1997) or has a temporary reduction in a threshold value (Schacter & Badgaiyan, 2001). This residual activity allows participants to make a faster and more accurate response the second time the stimulus is presented.

Repetition priming has been used extensively and successfully in visual identification research (see for example, Biederman & E. E. Cooper, 1991; Biederman & E. E. Cooper, 1992; Biederman, E. E. Cooper, & Hummel, 1992; Brooks, Rosielle & E. E. Cooper, 2002; Bruce & Young, 1986; A. W. Ellis, Flude, Young, & Burton, 1996; Fiser & Biederman, 1995; Fiser & Biederman, 2001; Hummel, 2001; Stankiewicz, Hummel, & E. E. Cooper, 1998). By changing the characteristics of the stimulus the second time the stimulus is presented (e.g., changing the color, the size, the mode of presentation), researchers can determine whether the alteration causes a reduction in priming (relative to the priming produced by presenting an identical version of the stimulus instead). No reduction in priming implies that the same representation was activated by both presentations of the

stimulus (despite the change in characteristics from the first to the second presentation of the stimulus), whereas a reduction in priming implies that different representations became active, or that it takes longer to activate the representation because of this change (see for example, Tenpenny, 1995). Thus the nature of the memory representation of a type of stimulus can be investigated by looking at what types of changes to a stimulus do and do not cause a reduction in priming.

Although there are a number of possible methods of investigating the memory representation used to accomplish visual identification, many researchers have argued that repetition priming has advantages over other methods (Biederman & E. E. Cooper, 1991, 1992; L. A. Cooper & Schacter, 1992; Roediger & Srinivas, 1993; Srinivas, Breedin, Coslett & Saffran, 1997; Tulving & Schacter, 1990). For example, same-different matching (are these two stimuli the same or different?) may be contaminated by short-term or working memory representations, which have different characteristics than the representations used for identification (Biederman & E. E. Cooper, 1992; R. Ellis & Allport, 1986; Rosielle & E. E. Cooper, 2001). Similarly, tasks that require participants to name a single presentation of a stimulus (such as naming rotated objects) may produce results contaminated by semantic and conceptual information (L. A. Cooper & Schacter, 1992). Repetition priming, however, is thought to provide a relatively pure measure of visual memory (Biederman & E. E. Cooper, 1992; L. A. Cooper & Schacter, 1992).

A consistent pattern of invariance in the priming of object identification to a wide variety of metric transformations (size, position, mirror image reflection, and rotation in depth) has been demonstrated by various researchers (Biederman & E. E. Cooper, 1991, 1992; Biederman & Gerhardstein, 1993, Fiser & Biederman, 1995; Fiser & Biederman,

2001; Hummel, 2001; Stankiewicz, et al., 1998). For example, Biederman and E. E. Cooper (1991, 1992) used repetition priming to investigate whether the representation used for object identification includes information concerning the size of an object and its position in the visual field. In the first block of trials, participants named line-drawings of common objects. Then in the second block of trials, participants named the same objects they saw in the first block, but sometimes the objects were presented at a different size or sometimes the objects were presented in a different position on the screen. Biederman and E. E. Cooper found no reduction in priming for size and position changed objects compared to objects that were identical in size to those viewed in the first block. They interpreted these results as evidence that the same memory representation was activated for the objects regardless of their size or position in the visual field, and thus concluded that the memory representation used for object identification was stored independently of the position or size of the object.

These results have important implications for theories of object identification. Most theories of object identification posit that size and position information are included in the representation used to identify objects (e.g., Tarr, 1989; Ullman, 1989, 1996) and thus were inconsistent with the results. A few theories, however, (e.g., Biederman, 1987; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996; Olhausen, Anderson, & Van Essen, 1993) do posit that objects are stored independently of position and size and thus were supported by the above results. Although the method of achieving size and position invariance differ vastly among these theories (e.g., compare Hummel & Biederman, 1992, and Olhausen, et al,, 1993), the empirical data suggest that a size and position invariant representation of some sort must be used during object identification.

If scenes and objects do share a common representational code, then the representation used for scene identification should also be coded independently of size and position information. In addition, there are other computational advantages to coding a scene independent of its size and position. Given that scenes can be viewed from many different viewpoints, information concerning the size of the scene is irrelevant for identification. For example, a "kitchen" is still a "kitchen" regardless of how much of the visual field the scene occupies. Thus, it would seem computationally efficient to ignore information concerning the size of the scene because this information is not used for identification. Further, a scene's (or portion of a scene's) position in the visual field can vary dramatically as one moves through the scene. Thus, it would seem to be computationally efficient to ignore position information for purposes of identification because this information does not seem necessary for identification.

However, there is evidence that size and position information may be extracted from scenes under certain conditions. Several studies indicate that metric position information is included in peoples' memory representations for maps (Evans & Pezdek, 1980; Presseon & Hazelrigg, 1984; Presson, DeLange, & Hazelrigg, 1989; Roskos-Ewoldsen, McNamara, Shelton, & Carr, 1998; Thorndyke & Hayes-Roth, 1982). Likewise, other studies have indicated that the representation used for navigation is, under certain conditions, sensitive to changes in the absolute position of the elements of the route. Some studies have found evidence that the precise metric positions of the elements of a scene are coded in memory during route learning (e.g., Christou & Bulthoff, 1999; Evans & Pezdek, 1980; Janzen, Schade, Katz, & Hermann, 2001; Pezdek & Evans, 1979; Wang, 1999), whereas other studies have found situations in which only the relative positions of the elements of the

scene are coded (e.g., Presson & Hazelrigg, 1984; Presson, DeLange, & Hazelrigg, 1989). For example, learning a route from a map tends to lead to metric coding of position whereas learning a route by walking through the route tends to lead to only relative coding of position (Allen & Willenborg, 1998; Presson & Hazelrigg, 1984). If the same representation used for scene identification is also used for scene navigation, then it could be the case that size and position information are included in the representation used to identify scenes. Further, some positions of objects in scenes are impossible (e.g., a lamp floating in mid air) and objects in a scene in impossible positions interfere with the identification of those objects (Biederman, Mezzanotte, & Rabinowitz, 1982). Thus, some types of position information may be included in the memory representation of a scene.

However, none of these studies directly tested scene identification, and there are reasons to suppose that the representation used for identification is different from other types of visual representations. For example, identification is thought to take place in the ventral cortical pathway, whereas navigation is thought to be controlled by the dorsal cortical pathway (Milner & Goodale, 1995; Mishkin, Ungerleider, & Macko, 1983). It could thus be the case that the representation used for scene identification is not the same representation used for scene navigation. Wang (1999) argued that although scene navigation requires metric information (such as size and position), representation of scene shape does not. If Wang is correct, then perhaps the representation used to identify scenes is coded independently of size and position information.

Episodic Recognition

Related to people's ability to identify scenes as a certain type (a "bedroom" or a "park") is the ability to recognize locations as a specific place one has been previously (e.g.,

"my bedroom" or "Hyde Park"). Not only is it important for people to extract the general identity of a scene, but it is also important to assess whether a particular location is familiar or unfamiliar. For example, navigating a previously experienced but largely unfamiliar environment (such as finding one's way around a new city during the first few weeks) requires one to not only identify general locations (e.g., "street", "park") but also to assess whether these locations are familiar or not. The latter task largely involves episodic memory retrieval (e.g., "Have I been here before or am I in a new location?").

According to Tulving and Schacter (1990), the processes that underlie perceptual (e.g., visual) priming are distinct from the processes that underlie episodic memory. Tulving and Schacter cited several lines of evidence that lead to this conclusion. For example, patients with amnesia have compromised episodic memory (they fail to remember events) but still show perceptual priming effects. Further, normal participants show several dissociations between episodic memory and priming effects. Whereas elaborating on the meaning of study material affects episodic memory, it has little effect on priming. Conversely, changing the modality of the study material affects priming but has little effect on episodic memory (see Tulving & Schacter, 1990, for a review).

In addition, the brain regions responsible for episodic memory seem to be separate from the brain regions responsible for other sorts of visual tasks. In a comprehensive review of 275 positron emission tomography (PET) and fMRI studies, Cabeza and Nyberg (2000) found areas in the prefrontal lobe and medial temporal lobe that become active during episodic memory retrieval that did not become active during other tasks, such as visual imagery.

Further, studies have found that the memory representation used for episodic memory is very different from the representation used for object identification (as measured through visual priming). In an additional set of experiments, Biederman and E. E. Cooper (1992) tested episodic recognition instead of identification after size transformations. After naming objects in a first block of trials, participants saw a second block of trials during which they indicated whether an object was seen in the first block, or whether the object was new. Sometimes the object in the second block was a different size than in the first block and sometimes the object was the same size as in the first block. Unlike object identification, Biederman and E. E. Cooper (1992) found longer response times and more errors when the object changed size from the first to second block compared to when the object was presented at an identical size in both blocks. They interpreted these results as evidence that the representation used for episodic recognition, unlike the representation used for object identification, includes information about the size of the object. A similar conclusion has also been reached by L. A. Cooper and Schacter (1992). Thus, the characteristics of the representation used for object identification, as measured by priming, may have very different characteristics than the representation used for episodic recognition.

Does the representation used for episodic recognition of scenes include information about the size and position of the scene? Some evidence suggests that position information might be included in the episodic representation for scenes (Diwadkar & McNamara, 1997; Nakatani, Pollatsek, & Johnson, 2002; Shelton & McNamara, 1997, 2001). For example, Diwadkar and McNamara (1997) had participants study scenes composed of collections of objects. Later, the participants made old/new judgments to the scenes. In some of the scenes, the objects had all changed position. In general, Diwadkar and McNamara (1997)

found a linear relationship between the participant's time to decide whether a scene was old or new and the amount of disparity between the positions of the objects in the study scene and the test scene. They interpreted these results as evidence that participants were coding in memory the precise positions of the objects in the scene.

However, these results are again problematic. The stimuli were arrays of objects arranged on a circular board. Like other studies that use abstract stimuli, it is unclear the extent to which these stimuli constitute "scenes" (see Epstein & Kanwisher, 1998; Henderson & Hollingworth, 1999). Further, a "different" scene in these experiments was a scene with the same objects present, but with one of those object in a different location (relative to the other objects). These distractors may have encouraged participants to simply pick a benchmark object in the scene and memorize the other objects relative to it (e.g., "hammer next to pencil next to keys next to book"). The response time function may have thus been a product of the time taken for participants to locate the displaced benchmark object in the scene.

Laterality

Determining whether there is hemispheric specialization for a task is one of the first steps in locating the cortical locus of a particular process. One way to investigate hemispheric specialization for visual identification is to present stimuli either to the left visual field or right visual field. Because the brain is cross-wired, stimuli presented to the left visual field are initially processed by the right hemisphere, whereas stimuli presented to the right visual field are initially processed by the left hemisphere (Springer & Deutsch, 1993). If one hemisphere is specialized to process a certain type of stimuli, faster response latencies for stimuli initially presented to that hemisphere are often found. For example,

face recognition often shows a left visual field (right hemisphere) advantage (e.g., De Renzi, Perani, Carlesimo, Silveri & Fazio, 1994; Levy, Trevarthan, & Sperry, 1972; Sergent, Ohta & Mac Donald, 1992; Sergent & Sinnoret, 1992), whereas word recognition often shows a right visual field (left hemisphere) advantage (Barton, Goodglass, & Shai, 1965; Mishkin & Forgays, 1952). In contrast, objects often show no hemispheric advantage or show inconsistent results (Biederman & E. E. Cooper, 1991b; Bryden & Rainey, 1963; Levine & Banich, 1982; McKeever & Jackson, 1979).

There are two basic interpretations of these hemispheric asymmetries (see Banich, 1997; Springer & Deutsch, 1993, for overviews). One interpretation is that the hemisphere that initially receives the information will always processes the information regardless of whether that hemisphere is specialized for that type of information or not. The faster response times for information presented to certain hemispheres thus reflect that hemisphere's advantage for processing that type of information. A second interpretation is that when a hemisphere receives information that it is not specialized to process it transfers that information to the other, more specialized hemisphere. The faster response times for information presented to certain hemispheres thus reflects the fact that the information does not have to travel to the other hemisphere in order to be processed. Regardless of the interpretation, hemispheric asymmetries are thought to reflect specialized processing.

Very little is currently known about the cortical location of scene identification although the results of Epstein and Kanwisher (1998) point to the right hemisphere. Epstein and Kanwisher (1998) found a larger degree of cortical activity in the right parahippocampal region than the left parahippocampal region in response to passively viewed scenes. Similarly, Spiers, et al. (2001) found that patients with unilateral damage to the right

temporal lobe were impaired (relative to patients with unilateral left temporal lobe damage) at recognizing locations in a virtual town, again implicating the right hemisphere in scene processing (and perhaps identification). However, because neither of these studies tested scene identification directly, it is difficult to know if these results can be generalized to identification.

Other studies suggest that the right hemisphere may be particularly well suited for scene identification. Hellige (1980) argued that the right hemisphere is specialized for early visual processing (such as extracting the global shape of the stimulus), which is thought to play a large role in scene perception (Oliva & Schyns, 1997; Schyns & Oliva, 1994). Likewise, the right hemisphere is also specialized for processing low spatial frequency information (Sergent, 1983), which is also thought to play a critical role in scene identification (Oliva & Schyns, 1997; Schyns & Oliva, 1994). Further, the right hemisphere is specialized for the processing of the global characteristics of a stimulus (as opposed to the local details) (Delis, Robertson, & Efron, 1986), which are also thought to play a role in scene identification (Loftus, Nelson, & Kallman, 1983). Thus, the right hemisphere appears to be a good candidate for the site of scene identification.

However, other evidence from the perception of paintings speaks against the role of the right hemisphere in scene perception. Zaidel and Kasher (1989) presented naturalistic and abstract paintings first in central vision and then (sometime later) either to the left or right visual field. Zaidel and Kasher found that participants had better memory for the abstract paintings presented in the right visual field (left hemisphere) but found no hemispheric advantage for the naturalistic paintings. Further, Bhattacharya and Petsche (2002), using EEG technology, found greater involvement of the right hemisphere when

viewing paintings, but only for experienced artists. The results call into question the right hemisphere's role in everyday, naturalistic scene identification.

Further evidence against the role of the right hemisphere as a candidate for scene identification comes from studies that investigate the type of information that each hemisphere is specialized to processes. Numerous studies point to the role of the left hemisphere in processing categorical relations (e.g., "above" or "next to") both within an individual object and between multiple objects (Hellige & Mitchimata, 1989; Kosslyn, Chabris, Marsolek, & Koenig, 1992; Kosslyn, et al., 1989). Generally, the right hemisphere is thought to processes coordinate relations whereas the left hemisphere is thought to process categorical relations. Given that the role of categorical relations in scene perception has been documented (see Rosielle, Crabb, & E. E. Cooper, 2002), it is thus possible that the left hemisphere is responsible for scene identification.

Overview of the Experiments

The question proposed by this research is whether scene identification, like object identification, is invariant to changes in position and size and whether the episodic recognition of scenes is not invariant to changes in position and size. The outcome of this research has implications for both scene identification and object identification. Few studies to date have investigated the memory representation that allows for the identification of scenes, and no study has investigated the role of metric information (i.e., size and position) in scene identification. Further, this research provides the first empirical test of Biederman's proposal that scenes and objects share a common representational code. If Biederman is correct, then scenes and objects should show similar patterns of priming after metric transformations. If scenes and objects should show different patterns of priming, then

Biederman is incorrect. Finally, these studies will bring together the literature on scene perception and object identification. By replicating experiments originally performed on objects with scenes, direct comparisons can be made between the two sets of results.

The following is an overview of these experiments. The purpose of this section is to explain the goals of each experiment. Later the specific methodology will be discussed in detail.

Experiment 1: Is Scene Identification Invariant to Size Transformations?

Experiment 1 tests whether the representation used for scene identification, like the representation used for object identification, is invariant to changes in size. In Experiment 1, participants identify scenes in two blocks of trials; sometimes in the second block the scene is a different size than the scene in the first block. The critical question is whether changing the size of the scene affects the participants' response to the scene relative to their response to a scene presented at an identical size.

Experiment 2: Is Scene Identification Invariant to Position Transformations?

Experiment 2 tests whether the representation used for scene identification, like the representation used for object identification, is invariant to changes in position. Experiment 2 is similar to Experiment 1, except that the scenes change position (relative to fixation) instead of changing size.

Experiments 3: Is Episodic Recognition of Scenes Invariant to Size Transformations?

Experiments 3 tests whether the episodic recognition of scenes (e.g., "have you seen this kitchen before?") is invariant to size transformations. The first block of trials is similar to Experiment 1. In the second block of trials, however, the participants is asked to indicate whether a scene is one they saw in the first block, while ignoring any changes in size.

Experiment 4: Is Episodic Recognition of Scenes Invariant to Position Transformations?

Experiment 4 tests whether the episodic recognition of scenes is invariant to changes in position. Experiment 4 is thus similar to Experiment 3, except that the scenes will vary in their position on the screen instead of their size.

Experiment 5: Does Scene Identification show Hemispheric Specialization?

The position manipulation in Experiments 2 and 4 (presenting the scenes to the left or right of fixation) allow for a convenient opportunity to investigate whether scene identification shows a hemispheric advantage. The laterality effects of scene identification is investigated in Experiment 5 by examining the first block data from Experiments 2 and 4, in which scenes were presented to either the left or right visual field. An advantage for either position (left or right) is evidence for the laterality of scene identification.

Methodological Considerations

Visual versus Semantic Priming

Of considerable importance in visual repetition priming experiments is whether the priming is due to priming of the visual representation or priming of other, non-visual, characteristics of the stimulus. As noted by several authors (e.g., Biederman & E. E. Cooper, 1992; Johnston & Dark, 1986), viewing a stimulus will activate multiple representations for that stimulus. For example, viewing a picture of a kitchen will not only activate a visual representation of the kitchen, but also, for example, semantic information associated with the kitchen and the name "kitchen" in the lexicon. Thus, any priming

associated with the picture of the kitchen could be due to the priming of the visual representation, the priming of semantic representations, or the priming of name representations. Further, because the participants are typically required to perform a task twice, improvement from the first to second presentation can also be due to general practice effects (e.g., increased familiarity with the task, greater confidence with the response procedures). How, then, can one be sure that any priming is due to the visual aspects of the stimulus and make claims regarding the nature of the visual representation?

The way to control for practice effects and semantic/conceptual priming is to include a condition in the second block of trials in which a stimulus is presented with the same name but different visual characteristics as a stimulus in the first block (i.e., a different exemplar) (Biederman & E. E. Cooper, 1991). If the priming observed is completely due to semantic or conceptual priming (and lacks a visual component), then there should be no difference in the priming of the same exemplar images and the priming of different exemplar images. The two different exemplar stimuli share different visual features but similar conceptual and semantic features; therefore, if only semantic/conceptual priming occurs, the stimuli should prime each other just as much as they should prime themselves. However, if there is a reduction in priming for different exemplars (which share the same conceptual features but different visual features) compared to same exemplars then it can be concluded that a portion of the priming is visual. Because the difference between the same and different exemplars is visual, any advantage in priming that the same exemplar trials have over the different exemplar trials represents the priming of the visual representation.

Thus, for the experiments testing the priming of scene identification (Experiments 1 and 2) a different exemplar condition will be included to determine if a portion of the

priming is visual. In the second block of trials, a scene will sometimes be presented that shares the same name but few visual characteristics as a scene presented in the first block (e.g., two pictures of different kitchens). If some of the priming is visual, then these two different exemplars of the same scene should not prime each other as much as two pictures that are identical in every way.

The Scene Identification Task

Also of importance is the choice of an appropriate scene identification task. Although the most obvious task is to have participants generate the name of a scene, this approach is problematic. Requiring participants to generate the name of other types of stimuli such as famous faces typically results in a large amount of variance in the data and can often be quite difficult (Cohen, 1990; Cohen & Faulkner, 1986; Young, Hay, & A. W. Ellis, 1985). Although scene naming was used successfully by Oliva and Schyns (1997), participants were only required to repeatedly generate the name of only a few different scenes (e.g., "city" and "highway") over the course of the experiment.

A way of avoiding the problems associated with naming is to use a matching task in which the picture of a scene is followed by a written name of a scene. Participants would be required to indicate with a button press whether the name of the scene matched the picture of the scene. Potter (1976) used a similar technique of matching names of scenes to pictures of scenes by presenting a name of a target scene before a rapid serial visual presentation sequence of scenes. Further, scene-name matching is likely analogous to the task used by Oliva and Schyns (1997) because in their experiments participants likely were aware of the names of all possible scenes during the experiment (because they had to repeatedly name one of four possible scenes). Thus, participants had to match a picture of a scene with a

memorized name of a scene. However, the current research will present the name of the scene after the picture of the scene so as not to prime the participant as to what to expect when viewing the picture (see also Brooks, Rosielle, & E. E. Cooper, 2002, who used this same procedure with faces).

Alternate Strategies

Also of importance in the current set of experiments is the possibility that participants may adopt a strategy of only encoding a few important objects in the scene and ignoring the rest of the scene. Given that the priming of individual objects is invariant to changes in size (Biederman & E. E. Cooper, 1992), if participants were to only encode a few discrete objects, then there would be little reason to suspect that the priming of scene identification would reflect anything other than the priming of multiple objects in the context of a scene. The priming of multiple objects under position and size changes may in itself be theoretically interesting as the visual system may treat two or more objects simultaneously presented differently than a single object in isolation (see Saiki & Hummel, 1996, for example) because the visual system may compute the spatial relations among the various objects as well as the objects themselves.

However, there are good reasons to suspect that in the current experiments participants will not (or cannot) use the strategy of searching for a few diagnostic objects in order to identify the scene. First, the literature on scene perception (discussed previously) suggests that scene identification probably does not occur by first recognizing the constituent objects. Second, the scenes in the following experiments will be presented at pre-saccade

levels (i.e., 210 ms[1].). Previous studies on scene perception suggest that scenes can be identified down to exposure durations of 45 ms (Oliva & Schyns, 1997; Schyns & Oliva, 1994). Given that it takes approximately 250 ms for the eyes to saccade, it seems unlikely, if not impossible, for participants to be able to scan the scene for a few diagnostic objects and base their conclusion of the scene's identity on this information because the exposure duration of the scene will not allow participants the time necessary to perform any visual scanning. Further, the name of the scene will be presented after the picture of the scene; thus participants will not be primed within a trial to what types of objects to look for in order to decide whether the name of the scene matches the picture of the scene. Finally, the names of the target scenes will not contain the name of any objects in the scene. For example, "bedroom" will not be used as a target scene because the name ("bedroom") will provide direct cues to the objects that may be present (a "bed"). Thus, participants will not be able to adopt a strategy of basing their response on whether the name of the scene contains the name of an object viewed in the scene.

[1] Pilot data indicate that an exposure duration of 210 ms is well suited for these experiments (given the visual angles of the scenes). At exposure durations of less than 210, the task becomes exceedingly difficult. At exposure durations of greater than 210, there is a possibility that participants will be able to scan the scene.

CHAPTER 2: IS SCENE IDENTIFICATION INVARIANCE TO CHANGES IN SIZE?

The goal of Experiment 1 was to test whether the representation used for scene identification is invariant to changes in size. A repetition priming paradigm was used in which participants were shown a picture of a scene followed by the name of scene. The participants had to decide whether the name of the scene matched the picture of the scene. On a second block of trials, for each scene type viewed in the first block, the participants were shown: the same scene at the same size they viewed in the first block, the same scene but at a different size, or a different photograph of a scene that shares the same name as a scene viewed in the first block (e.g., two pictures of different kitchens). See Figures 1 and 2 for examples of the experimental procedure and conditions. The participant's task in both blocks of trials was to determine whether the name of the scene matches the picture of the scene (Figure 2).

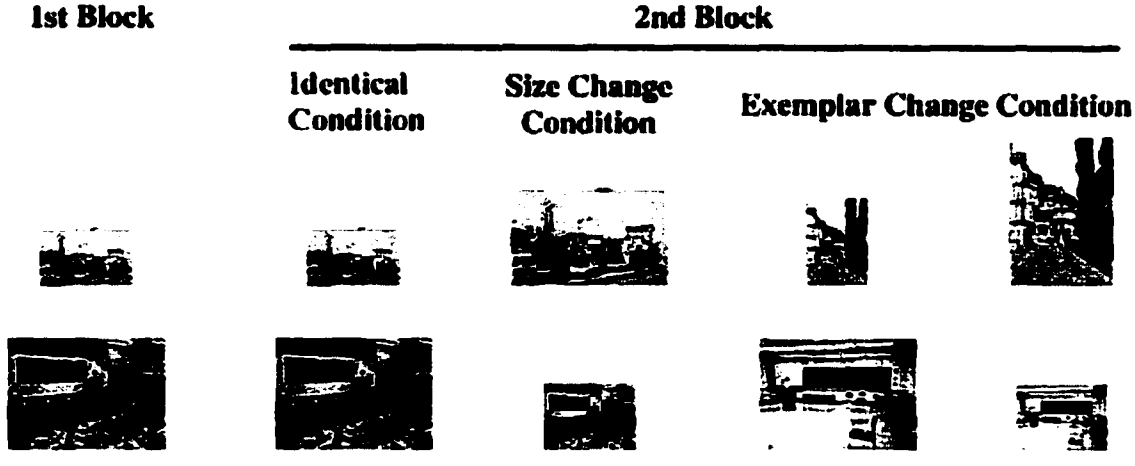

Figure 1. An illustration of the conditions in Experiment 1.

If the representation used for scene identification is invariant to changes in size, then changing the size of the scene between the first and second block should have no impact on the amount of priming observed. That is, a picture of a scene should prime a different sized version of the same scene just as much as it should prime an identical sized version of the scene because both sized versions of the scene should activate the same representation in memory. If, however, the representation used for scene identification is not invariant to changes in size then changing the size of the scene between the first and second block of trials should reduce the amount of priming observed. That is, a picture of a scene should prime a different sized version of the scene to a lesser degree than an identical sized version of the scene because a different representation in memory should be activated for different sized scenes.

Method

Participants

Participants were 32 college students from the research participant pool at Iowa State University. All participants were native English speakers, had normal or corrected-to-normal vision and all received course credit for participation.

Apparatus

Stimuli were presented on a 17 inch Apple color monitor with a resolution of 832 x 624 pixels and a vertical refresh rate of 75 Hz. Participants responded using a two button response box attached to a National Instruments NB-DIO-24 board that gives ±0.5 ms response time accuracy. The experiment was controlled by Picture Perception Lab Software (Kholmeyer, 1992) running on a Quadra 800 Macintosh computer.

The stimuli used in the experiment were forty-eight pairs of color photographs of real-world, readily nameable scenes (e.g., kitchen, office, street). Each pair of scenes was two scenes with the same name (e.g., "kitchen") but the pair were as different from each other as possible (e.g., different viewpoints, types of objects present, and color schemes). The large version of each scene was 512 x 512 pixels and subtended 6.2° x 6.2° of visual angle under the presentation conditions used in the experiment. The small version of each scene was 299 x 299 pixels and subtended 3.5° x 3.5° of visual angle. These particular dimensions were the same as those used by Biederman and E. E. Cooper (1992) in their object priming experiments.

The scene names were displayed in 72 point New Times Roman font. Under the viewing conditions used in the experiment, the names subtended no more than 6.2° of visual angle.

Procedure

Each participant saw two blocks of trials; a priming block (the first block of trials) and a test block (the second block of trials). On each trial during the experiment, each participant saw a fixation cue for 504 ms, followed by a picture of a scene for 210 ms, followed by a name for 1008 ms (Figure 2).

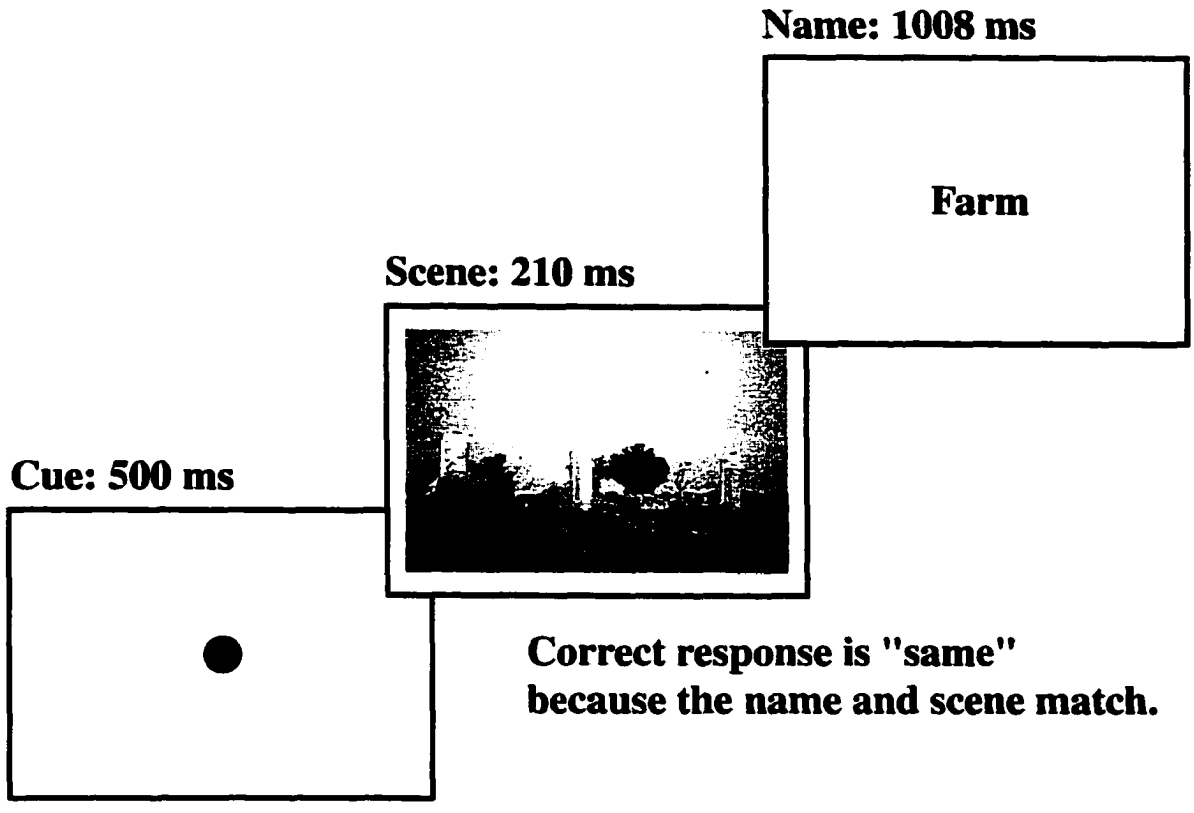

Figure 2. The trial sequence in Experiment 1. The size of the scene relative to the viewing area is not to scale.

The participant's task on both blocks of trials was to press the left button on the response box if the photograph of the scene matched the name of the scene and to press the right button on the response box if the name of the scene did not match the picture of the scene. Participants only used the fingers from their dominant hand to make their responses.

Before each block, participants were read a standard set of directions and were shown examples of small and large scenes. Participants were not told about the second block of trials until after the completion of the first block. Before the second block of trials, the participants was informed that some of the scenes in the second block will be the same as in the first block, but sometimes the scenes will change in size. Prior to the experiment,

each participant completed ten practice trials on which no data were collected. None of the scenes or names of scenes used in the practice trials was used in the experiment proper.

Each trial was self paced and participants initiated each trial sequence by pressing a mouse button in response to a prompt on the computer screen. Participants were told to respond as quickly as possible while striving for 90% accuracy. Response time and accuracy feedback were given to participants after each trial, and a participant's mean response time and error percentage were given at the end of each block.

Each participant completed 128 trials over the course of the experiment (64 trials per block). Half of the 128 trials (64 trials total; 32 trials per block) were distractor trials in which the name of the scene did not match the picture of the scene. The remaining 64 trials (32 trials per block) were target trials in which the name of the scene matched the picture of the scene. On each block, half of the scenes were the small version of a scene and half of the scenes were the large version of the scene. For the 32 target trials in the second block, the scene was a different size than in the first block 1/4th of the time (8 trials), was a different exemplar than in the first block 1/4th of the time (8 trials), was both a different size and a different exemplar than in the first block 1/4 of the time (8 trials), or was identical in every way to the scene viewed in the first block 1/4th of the time (8 trials). Across participants, each target scene appeared equally often in all conditions and each of the four versions of a scene (2 sizes x 2 exemplars) appeared equally often in the first block and second block of trials. Further, the order in which the scenes appeared was randomly generated for both the first and second block of trials.

For the distractor trials, half of the scenes were small and half of the scenes were large (for both the first and second block of trials). The presentation of the distractors in the

second block of trials mirrored the presentation of the targets; 1/4th of the time the distractor scene was a different size than in the first block, 1/4th of the time the distractor scene was a different picture of a distractor viewed in the first block, 1/4th of the time the distractor scene was a different size and exemplar than a distractor viewed in the first block, and 1/4th of the time the distractor was identical in every way to a distractor viewed in the first block. All participants saw the same set of distractor trials. When selecting names for the distractor trials, pictures of indoor scenes were paired with indoor names and pictures of outdoor scenes were paired with outdoor names. This paring was done to help avoid the possibility that a single salient feature would reveal that the name and the scene did not match (e.g., the name "basement" paired with any picture containing a sky would obviously be a mismatch).

Results

Prior to the collection of any data, a response criterion of 70% correct was established. Those participants who failed to reach this criterion had their data replaced by data from new participants. Using this criterion, data from four participants were replaced by data from new participants.

The response times and error rates for Experiment 1 can be seen in Figures 3 and 4. The first block of trials had a mean response time of 682 ms and a mean error rate of 4.5%. For the distractor trials, the mean response time was 698 ms and the mean error rate was 6.1%. The response time and error rate data from block 2 of Experiment 1 were analyzed with separate one-way within participants analyses of variance (ANOVAs) with Second Block Condition (identical scene, size change, and exemplar change) as the single independent variable.

The goal of this experiment was to measure whether scene identification is sensitive to changes in size. Misidentifications (i.e., errors) during the first block of trials may indicate (among other possibilities) that the wrong representation became active in response to the scene. Thus, one would not expect any priming to a second block scene that was misidentified in the first block, not because of any possible size changes but because the representation is initially activated (i.e., primed) only on second block and not the first. As such, only the data from scenes in the second block in which participants made a correct response in the first block were included in the response time and error rate ANOVAs. Likewise, for the response time analysis, only data from second block trials in which the participant made a correct response in the second block were included in the analysis.

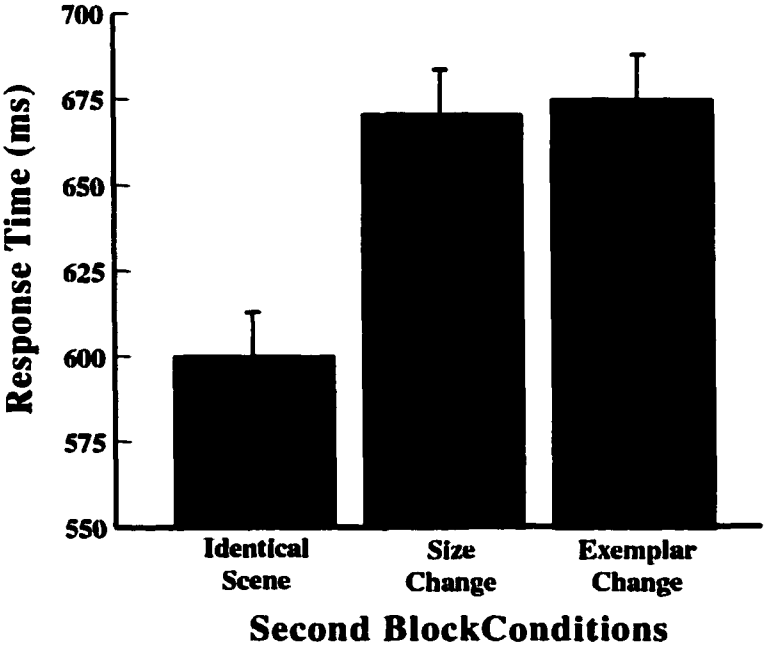

Figure 3. Mean response times for the second block conditions in Experiment 1. The error bars represent the standard error of the mean (Loftus & Loftus, 1988).

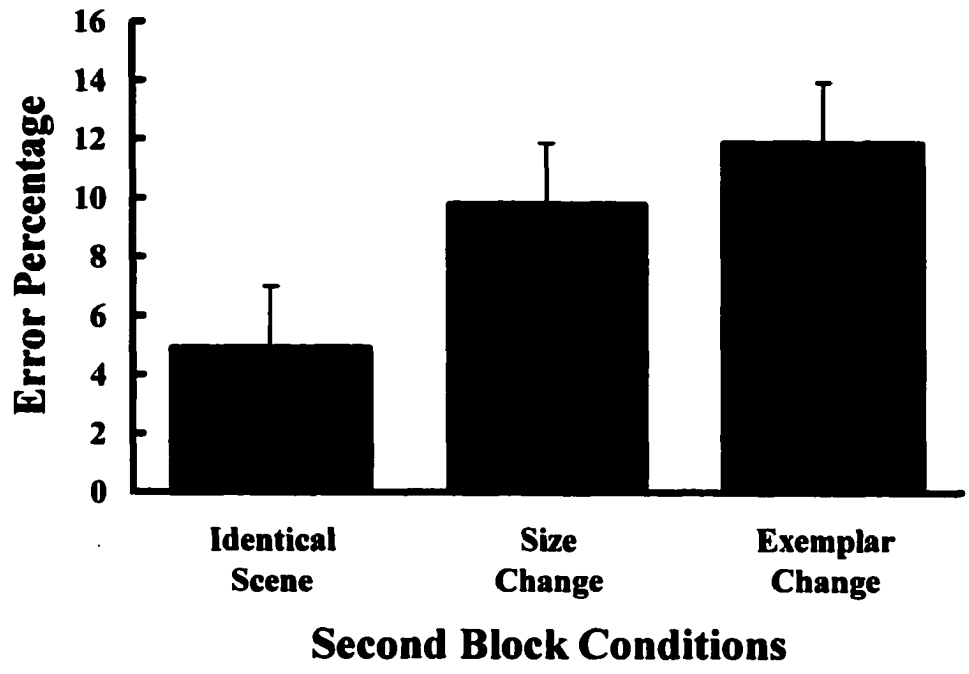

Figure 4. Mean error percentage for the second block conditions in Experiment 1. The error bars represent the standard error of the mean (Loftus & Loftus, 1988).

For the response times, the ANOVA revealed a reliable effect of Second Block Condition, $F(2,62) = 4.93$, $p < .02$, $MSE = 10281$. Post hoc tests using the least significant difference method (LSD) produced an LSD value of 50.67 ms. The LSD test revealed that the identical condition produced reliably faster response times than both the size change and exemplar change conditions. No other comparisons were statistically reliable.

A similar pattern of results was found for the error rate analysis. The ANOVA revealed a reliable effect of Second Block Condition, $F(2, 62) = 8.33$, $p < .001$, $MSE = .05$. For the post hoc tests, the value of LSD was 3.5%, indicating that the identical condition produced reliably fewer errors than the size change or exemplar change conditions. No other comparisons were statistically reliable.

A 2X2 factorial ANOVA was also conducted on the response time and error rate data from Experiment 1 with First Block Size (small versus large) and Second Block Size (small versus large) as the two factors. The results of this analysis can be seen in Figures 5 and 6.

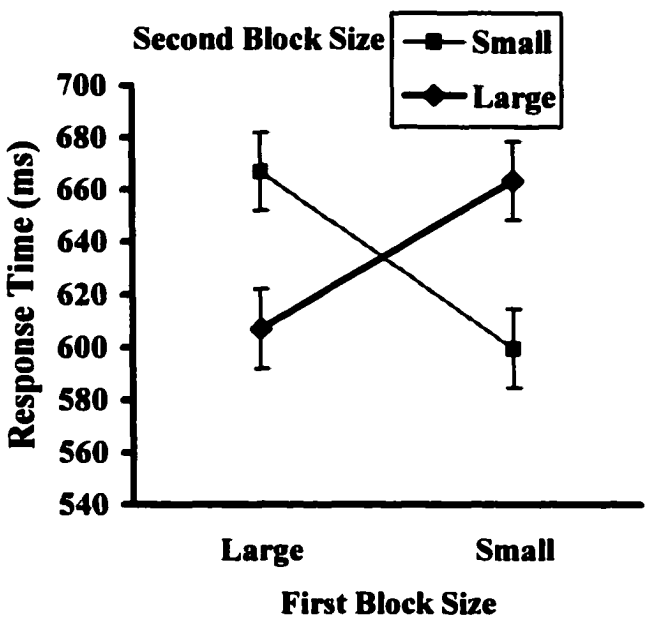

Figure 5. Mean response times as a function of size (small versus large) and block (first versus second) for Experiment 1. Error bars represent the standard error of the mean.

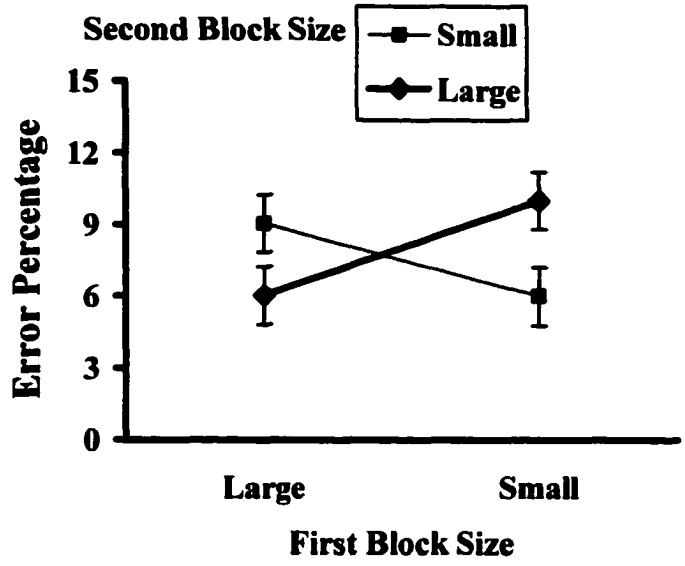

Figure 6. Mean error percentage as a function of size (small versus large) and block (first versus second) for Experiment 1. Error bars represent the standard error of the mean.

For the response times, the ANOVA revealed only a reliable interaction between First Block Size and Second Block Size, $F(1,31) = 288$, $p < .001$, $MSE = 420$. Likewise, for the error rate ANOVA, only the interaction was statistically reliable, $F(1,31) = 18.6$, $p < .001$, $MSE = .086$.

Discussion

Vital to the logic of Experiment 1 is that some of the priming observed in the identical condition must be visual in nature, otherwise no conclusions about the nature of the representation used to visually identify scenes could be drawn. If some of the priming was visual, there should be reliably faster response times and reliably fewer errors in the identical condition compared to the exemplar change condition. Such a difference was found for both response times and errors, indicating that at least a portion of the priming was visual (and not conceptual in nature).

The major result of the Experiment 1 was that reduced priming was found in the size change condition compared to the identical condition. The reduction in priming occurred for both response times and errors. This pattern of results indicates that the representation used to identify scenes is sensitive to changes in size.

These results do not support Biederman's (1988) hypothesis that scenes and objects share a common code. If Biederman were correct, then sized changed scenes should be responded to just as quickly and accurately as identical scenes, which is the same pattern observed for objects. However, this did not occur. In fact, Experiment 1 failed to find evidence of any visual priming whatsoever for size-changed scenes. For both response times and errors, scenes that were presented at a different size were responded to only as quickly and accurately as the exemplar changed scenes.

This pattern of results indicates a substantial difference between the representation used for object identification, and the representation used for scene identification. All experiments testing for metric invariance in object identification have found that the priming of objects is not sensitive to size transformations (Biederman & E. E. Cooper, 1992; Fiser & Biederman, 1995; Fiser & Biderman, 2001; Hummel, 2001). These results thus suggests that scenes and objects are represented differently in memory for the purposes of identification.

Further, these results indicate a major difference between the representation used for scene identification and the representation used for face identification. Like objects, faces show invariance to size transformations for priming (Brooks, Rosielle, & E. E. Cooper, 2002), despite the role that metric information (such as size) is thought to play in face identification (see for example E. E. Cooper & Wojan, 2000). Thus, the results of the

Experiment 1 suggest that scenes and faces are also represented very differently in memory for the purposes of identification. Scenes are the first example of a visual stimulus that uses a representation that is not size invariant.

The results of Experiment 1 establish scenes as class of stimuli different from faces and objects, neither of which show reduced priming after size transformations. The question remains, however, whether this difference is unique to size transformations or can occur with other metric transformation as well. Experiment 2 will attempt to answer this question by looking at a second metric transformation; position.

CHAPTER 3: IS SCENE IDENTIFICATION INVARIANT TO CHANGES IN POSITION?

The purpose of Experiment 2 was to test whether the memory representation used for scene identification is sensitive to changes in position in the visual field. The question of whether the representation used to identify scenes is invariant to position information is independent of the question of whether the representation used to identification scenes is invariant to changes in size. For example, it could be the case that position information is included in the representation whereas information concerning the size of the scene is not.

Experiment 2 employed the same general methodology as Experiment 1, but the scene's position on the computer screen varied (rather than its size). During the first block of trials, half of the scenes were presented to the left of fixation and half of the scenes were presented to the right of fixation. In the second block of trials, a scene was either presented in the same position that is was in the first block of trials (e.g., if a scene was presented to the left of fixation in the first block then it was presented to the left of fixation in the second block of trials) or in a different position than it was presented in the first block of trials (e.g., if a scene was presented to the left of fixation in the first block then it was presented to the right of fixation in the second block of trials) (Figure 7). As in Experiment 1, a different exemplar change condition was included to determine if some of the priming observed is visual (and not conceptual) in nature.

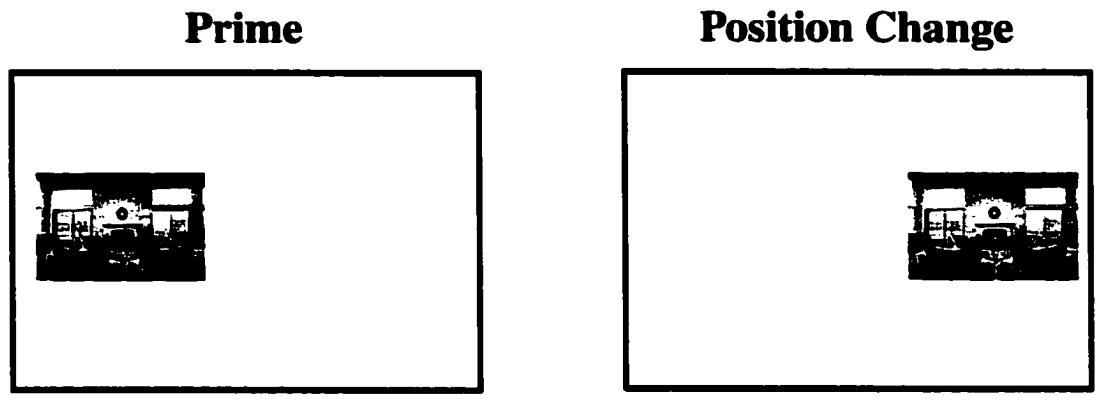

Figure 7. An illustration of the position change manipulation in Experiment 2. The size of the scene relative to the viewing area is not to scale.

The predictions for Experiment 2 closely mirrored the predictions of Experiment 1. If the priming of scene identification is invariant to changes in position, then a picture of scene should prime the same scene presented in a different position just as much as it should prime the same scene presented in the same position (because the same memory representation will become active regardless of position). If, however, the priming of scene identification is not invariant to changes in position then a picture of a scene should not prime the same scene presented in a different position (because these two scenes will activate different memory representations) as much as it should prime the same scene presented in the same position.

Method

Unless otherwise noted, the apparatus for collecting data and the general procedure used in Experiment 2 was identical to that used in Experiment 1.

Participants

Participants were 48 college students from the research participant pool at Iowa State University. All participants were native English speakers, had normal or corrected-to-

normal vision, and received course credit for participation. None of the participants took part in any of the other experiments reported here.

Apparatus

The apparatus for controlling the experiment and collecting data was identical to that used in Experiment 1.

The small sized scenes used in Experiment 1 were used in Experiment 2. The small scenes were used to allow for the scene to undergo a greater degree of displacement on the screen than would the large scenes. Each scene was 299 x 299 pixels and subtended 3.5° x 3.5° of visual angle under the viewing conditions used in the experiment. Each scene was displayed 2.4° to the left or right of fixation, making the closest point of the scenes to fixation 0.65°

Procedure

The procedure for Experiment 2 was almost identical to the procedure used in Experiment 1. However, instead of varying the size of the scenes, the position of the scenes varied instead. Thus, in the identical condition, the scene was presented in the same location on the screen in the second block of trials as in the first block of trials. In the position change condition, the scene was presented in a different location on the screen across blocks (e.g., if the scene was presented to the left of the fixation cue in the first block then it would be presented to the right of the fixation cue in the second block in the position change condition). In the exemplar change condition, a different picture of a scene that shares the same name as a scene viewed in the first block was presented in the second block of trials.

Results

Like Experiment 1, a performance criterion of 70% correct was established. Those participants who failed to reach this criterion had their data replaced by data from new participants. Using this criterion, data from four participants were replaced by data from new participants. The first block of trials had a mean response time of 655 ms and a mean error rate of 7.1%. For the distractor trials, the mean response time was 572 ms and the mean error rate was 6.3%.

The response times and error rates for Experiment 2 can be seen in Figures 8 and 9. The response time and error rate data from Experiment 2 were analyzed with separate one-way within participants analyses of variance (ANOVAs) with Second Block Condition (identical scene, position change, and exemplar change) as the single independent variable. Only the data from scenes in the second block in which participants made a correct response in the first block were included in the response time and error rate ANOVA. Likewise, for the response time analysis, only data from second block trials in which the participant made a correct response in the second block were included in the analysis.

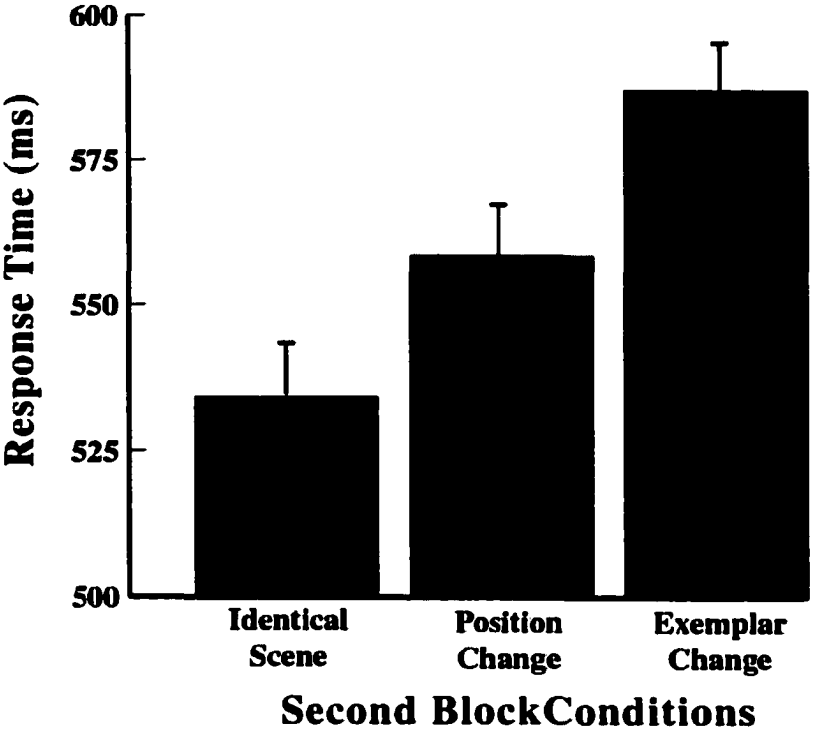

Figure 8. Mean response times from Experiment 2. The error bars represent the standard error of the mean.

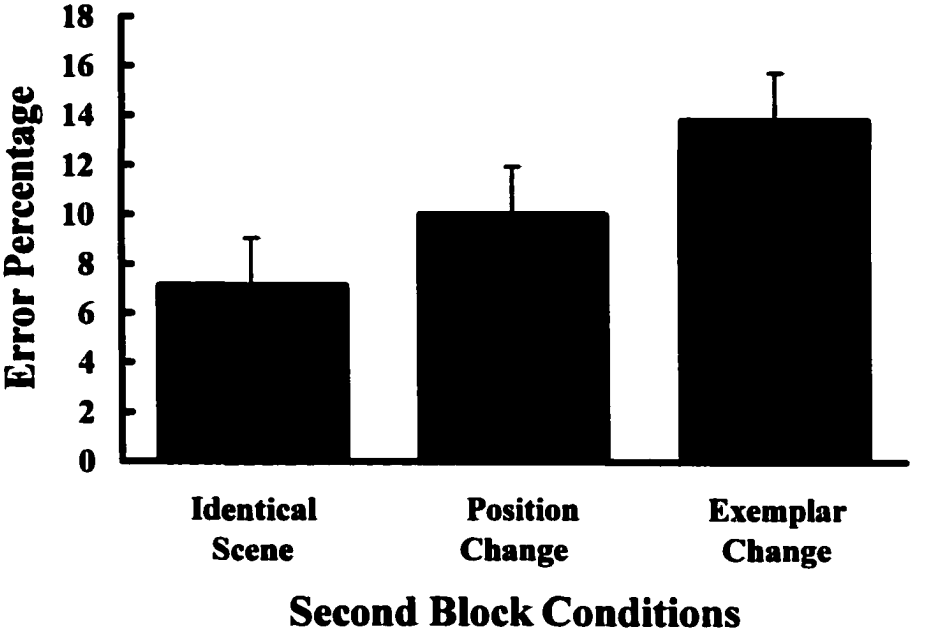

Figure 9. Mean error percentage for the conditions in Experiment 2. Error bars represent the standard error of the mean.

For the response time data, the ANOVA revealed a reliable effect of Second Block Condition, $F(2, 94) = 9.88$, $p < .001$, $MSE = 3495$. Post hoc tests using LSD produced an LSD value of 24 ms, indicating reliably faster response times in the identical condition than in both the exemplar change and position change condition, and reliably faster response times in the position change condition than in exemplar change condition.

For the error rate data, the ANOVA again revealed a reliable effect of Second Block Condition, $F(2, 94) = 6.02$, $p < .01$, $MSE = .09$. The value of LSD was 3.8%, indicating reliably fewer errors in the identical condition compared to the exemplar change condition. No other comparisons were statistically reliable.

A 2X2 factorial ANOVA was also conducted on the response time and error rate data from Experiment 2 with First Block Size (small versus large) and Second Block Size (small versus large) as the two factors. The results of this analysis can be seen in Figures 10 and 11.

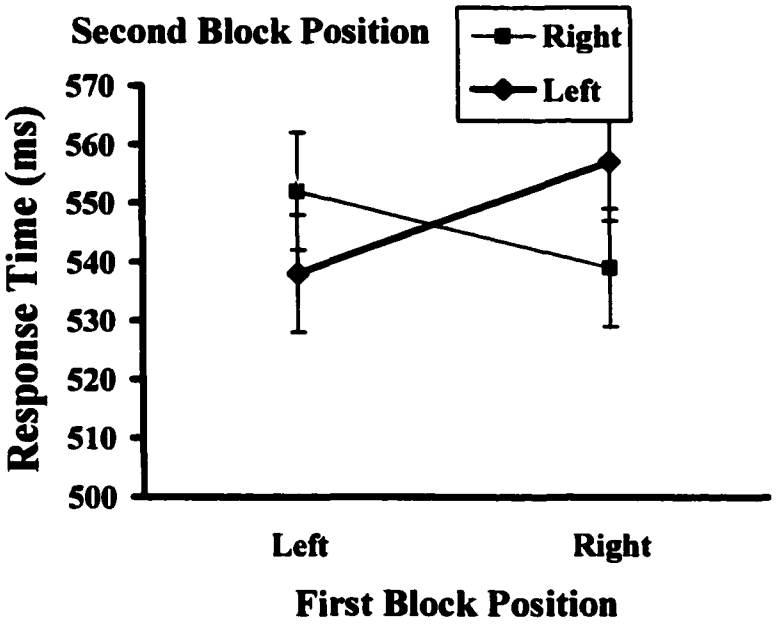

Figure 10. Mean response time as a function of block and position for Experiment 2. Error bars represent the standard error of the mean.

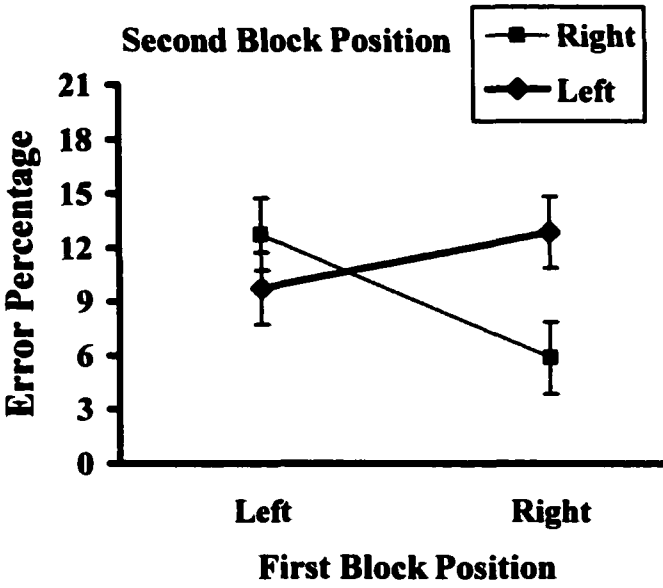

Figure 11. Mean error percentage as a function of block and position for Experiment 2. Error bars represent the standard error of the mean.

For the response times, the ANOVA revealed only a reliable interaction between First Block Size and Second Block Size, $F(1,47) = 4.34$ $p < .05$, $MSE = 20599$. Likewise, for the error rate ANOVA, only the interaction was statistically reliable, $F(1,47) = 19.6$, $p < .001$, $MSE = .18$.

Discussion

In general, the results of Experiment 2 mirror those of Experiment 1. Critically, the identical condition produced faster response times and fewer errors than the exemplar change condition, indicating that some of the priming was visual.

In addition, the identical condition produced faster response times than the position change condition, indicating that changing the position of the scene reduces the amount of priming. Thus, the lack of invariance to metric changes found in Experiment 1 is not limited to size transformations. For the response times, Experiment 2 found a lack of invariance to position changes as well. Although no reliable difference was found between the identical condition and the position change condition for the error data, note that the trend is the error data is the same as the reliable trend in the response time data.

This finding provides further evidence against Biederman's (1988) hypothesis. Biederman posited that position information, like size information, is not a part of the memory representation used to identify scenes. However, in Experiment 2, reduced priming was found when the scene changed position compared to when the scene was in an identical position, indicating that scene identification is sensitive to changes in position.

The results of Experiment 2 are consistent with the hypothesis that the representation used to identify scenes includes information about the position of the scene. Experiment 2 found evidence that position information is included in the representation used to identify

scenes, a result clearly in support of this idea. Thus, scene identification appears to code both size and position information for the purposes of identification.

One difference between Experiment 1 and Experiment 2 is that Experiment 2 found some evidence of visual priming after the position changes, whereas Experiment 1 found no evidence of visual priming after size changes. In Experiment 2, the position changed scenes were identified more quickly (but not more accurately) than the different exemplar scenes, indicating that some visual priming did occur after position changes. In contrast, no such difference between size and exemplar changed scenes was found in Experiment 1 (although a very small, but non-significant) trend was apparent.

There are several possible reasons for why this difference between the two experiments occurred. First, it could be that the Experiment 1 simply lacked the power to detect visual priming when the scene changed size. There was a small but non-significant advantage to identifying a size changed scene in Experiment 1 compared to a different exemplar, and this advantage occurred for both response times and errors.

Second, it could also be the case that scene identification depends more heavily on size information than it does position information, thus producing greater sensitivity to size changes. This greater sensitivity to size changes could be accomplished if the scenes were coded in memory using a very precise calculation of size, but a less precise calculation of position. Changing the size of the scene would thus be more disruptive to priming than changing the position of the scene, because the representation used for scene identification would tolerate some variance in position. This trend was observed in the data. Changing the size of the scene eliminated visual priming whereas changing the position of the scene did not. The scene recognition system may code scenes in this way because position

changes may be more frequent than size changes when interacting with a scene. For example, as I am sitting at my desk typing this, there is a noticeable displacement of the elements of the scene I am viewing caused by changes in my posture and head position, but no corresponding (noticeable) change in size. Thus, the scene recognition system may have developed a greater tolerance to position changes (than size changes) to compensate for a greater degree of natural variation in the input. Although this interpretation speaks to partial invariance to position changes for scene identification, it is still incompatible with Biederman's idea of complete position invariance.

CHAPTER 4: IS EPISODIC RECOGNITION OF SCENES INVARIANT TO CHANGES IN SIZE?

Experiments 1 and 2 found that the priming of scene identification is not invariant to changes in size and position. However, identifying a scene as a certain type (e.g., "kitchen") is not the only type of recognition that can occur with scenes. Episodic recognition, in which a specific scene is recognized as having been viewed in the past, can occur as well, and there are reasons to suppose that the representations underlying identification and episodic recognition are different. Thus, Experiments 3 tested whether the representation used for episodic recognition of scenes, like the representation used for the identification of scenes, is sensitive to changes in size by testing the episodic recognition for scenes after size transformations. Unless otherwise noted, the general procedures were conducted in the same way as in Experiments 1 and 2.

Method

Participants

Participants were 32 college students from the research participant pool at Iowa State University, selected with the same criteria as the other experiments. None of the participants took part in any of the other experiments reported here.

Apparatus

The apparatus for controlling the experiment and collecting data was identical to that used in Experiment 1. The same small and large scenes were used and were presented under the same viewing conditions as in Experiment 1.

Procedure

The first block of trials and the instructions given to the participant in Experiment 3 were identical to the first block of trials and instructions in Experiment 1. In the second block of trials, (64 trials) half of the scenes were "old" (viewed in the first block) and half of the scenes were "new" (not viewed in the first block). Of the "old" scenes, half were presented at the same size as in the first block and half were presented at a different size as in the first block. Only target scenes from the first block were used as "old" scenes in the second block. Each small and large version of the "old" scenes was presented an equal number of times across participants. All participants saw the same set of "new" scenes.

Unlike the first two experiments, no name was presented after the scenes in the second block; only the picture of the scene was presented. The participant's task in the second block was to press the left button on the response box if the scene was "old" and press the right button on the response box if the scene was "new". Participants were told to ignore any size changes that took place between the first and the second block; if a scene was the same as one they saw in the first block, despite any changes in size, the correct response was "old". Prior to the beginning of the second block, participants were read a set of directions and shown examples of scenes that were considered "old" and "new". Participants were not aware of the second block of trials (and thus the episodic memory task) until after the completion of the first block. The examples of "old" scenes shown to the participants were drawn from the practice set of scenes.

Results

The results of Experiment 3 can be seen in Figures 12 and 13. Like Experiments 1 and 2, a performance criterion of 70% was established prior to the beginning of the experiment. Eight participants failed to meet this criterion and had their data replaced.

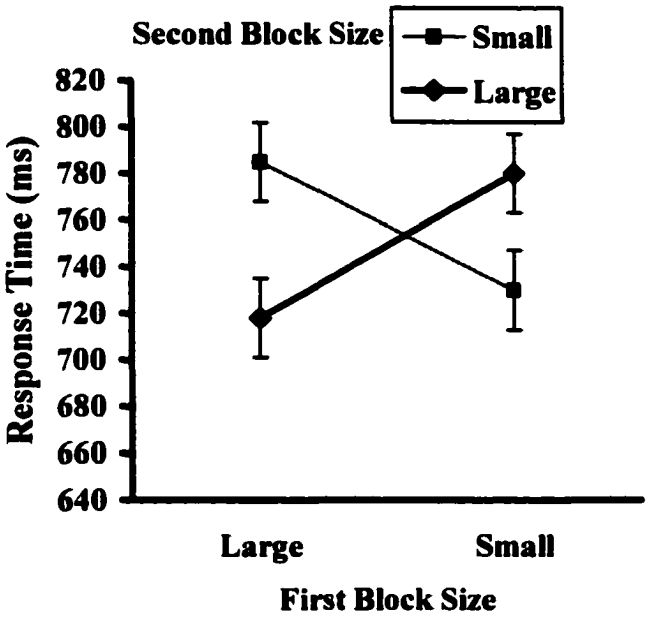

Figure 12. Mean response times as a function of size (small versus large) and block (first versus second) for Experiment 3. Error bars represent the standard error of the mean.

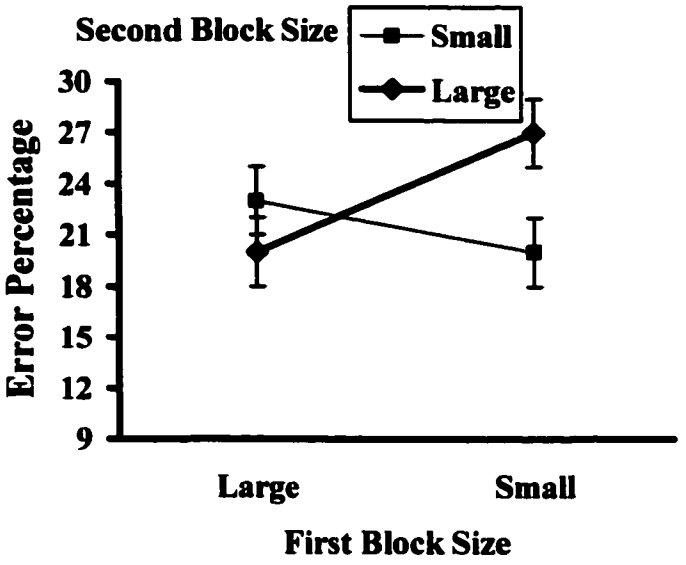

Figure 13. Mean error rates as a function of size and block for Experiment 3. Error bars represent the standard error of the mean.

The response time and error rate data from Experiments 3 were analyzed using separate 2X2 factorial ANOVAs with First Block Size (small or large) and Second Block Size (small or large) as the two factors. Only data from the old trials were included in the ANOVAs. For the response time analysis, a main effect of First Block Size was found, $F(1,31) = 4.53$, $p < .05$, $MSE = 531.68$, indicating that scenes that were initially presented as large were recognized more quickly than scenes than were initially presented as small. In addition, a reliable two-way interaction between First Block Size and Second Block Size was found, $F(1,31) = 11.63$, $p < .01$, $MSE = 9417$. For the error rate data, the only reliable effect was the interaction between First Block Size and Second Block Size, $F(1,31) = 6.80$, $p < .05$, $MSE = 1.15$. Thus, for both response times and errors, changing the size of the scene from the first block to the second block was detrimental to episodic recognition.

In addition, the discriminability index (d') was calculated in order to determine whether participants were able to discriminate old from new stimuli. The value of d' was 1.55. A test of a single sample mean found that this value was different from zero, $t(31) = 11.92$, $SE = .13$, $p < .001$, indicating that participants were indeed able to discriminate the old and new stimuli.

Discussion

The results of Experiment 3 indicate that the representation used for episodic recognition is sensitive to changes in size. Scenes that were viewed at a different size in the second block than in the first block (corresponding to the large-small and small-large conditions) took longer to identify and were more prone to errors than scenes that were viewed at the same size as in the first block.

Unlike Experiments 1 and 2, the results of Experiment 3 mirror the results found with the episodic recognition of objects. Objects show the same sensitivity to size changes for episodic recognition. Thus, it appears that the representation used for the episodic recognition of both objects and scenes includes information about the size of the stimulus.

Experiment 3 found that the representation used for the episodic recognition of scenes is sensitive to changes in size. Experiment 4 will examine whether the representation used for the episodic recognition of scenes is also sensitive to changes in position.

CHAPTER 5: IS EPISODIC RECOGNITION OF SCENES INVARIANT TO CHANGES IN POSITION?

The goal of Experiment 4 was to test whether the episodic recognition of scenes is sensitive to changes in position. Experiment 4 was thus similar to Experiment 3, except the scenes varied in position instead of size. Unless otherwise noted, the general procedure and the presentation conditions were identical to those used in Experiment 3.

Method

Participants

Participants were 48 from the research participant pool at Iowa State University, selected with the same criteria as the other experiments. None of the participants took part in any of the other experiments reported here.

Apparatus

The apparatus for controlling the experiment and collecting data was identical to that used in the other experiments. As in Experiment 2, only the small versions of the scenes were used in Experiment 4 to allow for the position manipulation to maximally displace the scene. The scenes in Experiment 4 were presented under the same presentation conditions used in Experiment 2.

Procedure

The first block of trials in Experiment 4 and the directions given to the participant were identical to those in Experiment 2. In the second block of trials, (64 trials) half of the scenes were "old" (viewed in the first block) and half of the scenes were "new" (not viewed in the first block). Of the "old" scenes, half were presented in the same position as they were presented in the first block and half were presented in a different position as in the first

block. Additionally, half of the "old" scenes appeared to the left of fixation and half appeared to the right. Only target scenes from the first block were used as "old" scenes in the second block. Across participants, each scene appeared equally often to the left of fixation and to the right of fixation. All participants saw the same set of "new" scenes.

As in Experiment 3, no name was presented after the scenes in the second block; only the picture of the scene was presented. The participant's task in the second block was to press the left button on the response box if the scene was "old" and press the right button on the response box if the scene was "new". Participants were told explicitly to ignore any position changes that took place between the first and the second block; if a scene was the same as one they saw in the first block, despite any changes in position, the correct response was "old". Prior to the beginning of the second block, participants were read a set of directions and shown examples of scenes that were considered "old" and "new". Like Experiment 3, participants were not aware of the second block of trials until after the completion of the first block. The examples of "old" scenes shown to the participants were drawn from the practice set of scenes.

Results

The results of Experiment 4 can be seen in Figures 14 and 15. As in all prior experiments, a performance criterion of 70% correct was established; five participants failed to meet this criterion and had their data replaced by data from new participants.

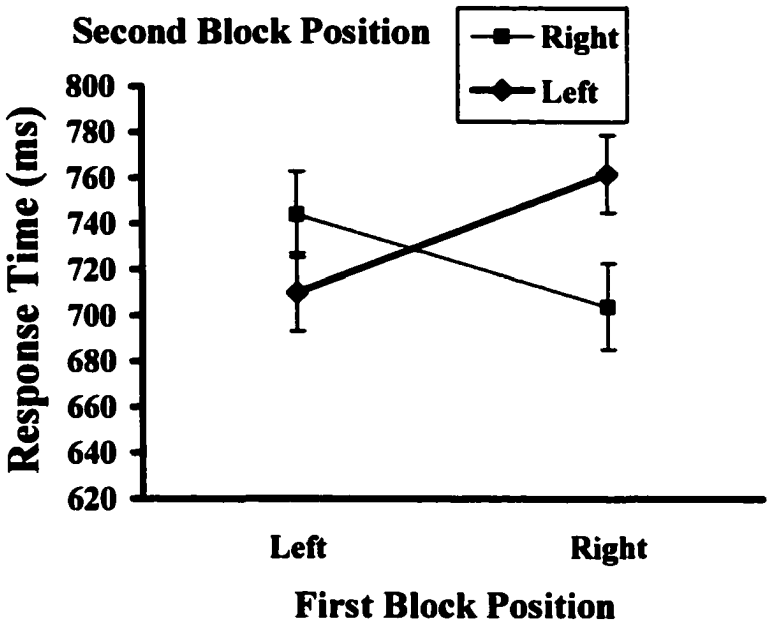

Figure 14. Mean response time as a function of block and position for Experiment 4. Error bars represent the standard error of the mean.

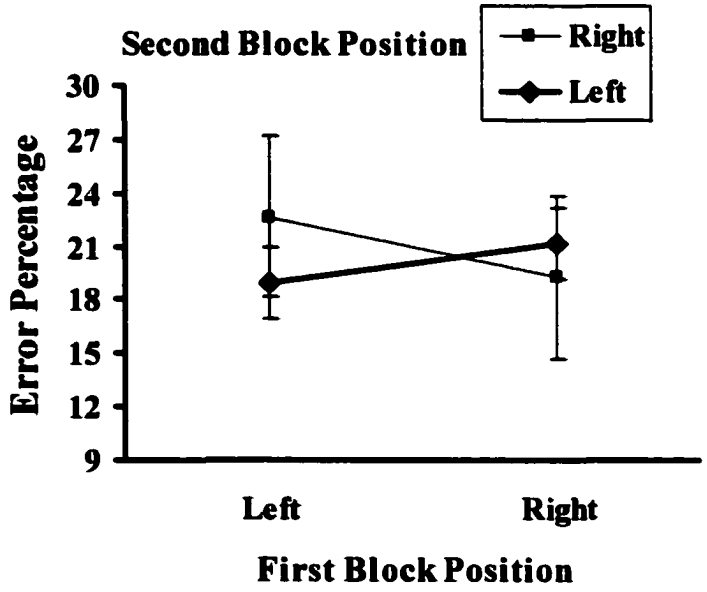

Figure 15. Mean error percentage as a function of block and position for Experiment 4. Error bars represent the standard error of the mean.

The response time and error rate data from Experiments 4 were analyzed using separate 2X2 factorial ANOVAs with First Block Size (small or large) and Second Block Size (small or large) as the two factors. Only data from the old trials were included in the ANOVAs. For the response times, the ANOVA revealed only a reliable interaction between First Block Size and Second Block Size, $F(1,47) = 6.08$, $p < .02$, $MSE = 16978$. For the error rate ANOVA, the interaction failed to reach significance, $F(1,47) = .145$, ns. However, the basic trend in the error rate data is the same as the significant trend in the response time data. In addition, the discriminability index (d') was calculated in order to determine whether participants were able to discriminate old from new stimuli. The value of d' was 2.05. A test of a single sample mean indicated that this value was different from zero, $t(47) = 20.50$, $SE = .10$, $p < .001$.

Discussion

The results of Experiment 4 indicate that episodic recognition of scenes is sensitive to changes in position. Participants were slower to determine that a scene was one they had seen previously when the scene was presented in a different position. These results suggest that the memory representation used for the episodic recognition of scenes includes information about the position of the scene.

It was somewhat surprising that participants found Experiments 3 and 4 so difficult. The overall error rate for the second block trials was 22.89% for Experiment 3 and 20.53% for Experiment 4. This result seems to be in contrast with previous studies that have found exceedingly good episodic memory for large numbers of scenes (e.g., Shepard, 1967; Standing, 1973). For example, Shepard (1967) had participants study 600 pictures from magazines and then later discriminate the studied pictures from new pictures. The error rate

for this task was only 1.5%. Similarly, Standing (1973) found only 17% errors when the number of pictures that were initially studied was increased to 10,000. Why, then, did the participants in Experiments 3 and 4 commit so many errors with only a relatively small number of scenes?

The answer may lie in the exposure duration of the scenes used in Experiments 3 and 4 (210 ms). Studies that have displayed scenes for prolonged exposure durations (e.g., Shepard, 1967; Standing, 1973) have found relatively good episodic memory for the scenes whereas studies that have used shorter exposure durations have found relatively poor episodic memory for scenes (Potter, 1976; Potter & Levy, 1969). For example, Potter and Levy (1969) (see also Potter, 1976) tested participants' ability to remember the masked presentation of sixteen pictures at varying exposure durations. Potter and Levy found that at 125 ms, participants could only identify 11% of the pictures they had seen previously. When the exposure duration of the scene was increased to 333 ms, the correct recognition increased to 45%, and at two-second exposure durations, participants performed close to 100%. Potter (1976) argued that although scenes can be "understood" (and presumably recognized) at low exposure durations, time is need to make the representation immune to masking from subsequent visual input. More recently, Rensink (2000) argued for a very similar mechanism; focused spatial attention (which takes time to apply) is necessary to stabilize the scene representation in order for a scene representation to enter long-term memory. Thus, the error rates found in Experiments 3 and 4 are compatible with prior research.

The representation used for the episodic recognition of scenes is not only sensitive to size changes, but to position changes as well. Next, the laterality of scene identification will be investigated.

CHAPTER 6: DOES SCENE RECOGNITION SHOW LATERALITY EFFECTS?

The goal of Experiment 5 was to test the laterality of scene identification. In Experiments 2 and 4, scenes were presented either to the left or right visual field. Stimuli presented to the left visual field are initially processed by the right hemisphere and stimuli presented to the right visual field are initially process by the left hemisphere. If either the left or right hemisphere is specialized for scene identification, then scenes presented to the contralateral visual field should be identified more quickly and accurately than scenes presented to the ipsilateral visual field.

The data from the first block of trials in Experiments 2 and 4 were analyzed in order to examine the laterality of scene identification. In the first block of Experiments 2 and 4, half of the scenes were presented to the left of fixation and half of the scenes were presented to the right of fixation. Further, each scene appeared equally often to the left and right of fixation over the course of the first blocks in Experiments 2 and 4. If the cortical locus of scene identification is in the right hemisphere, then scenes presented to the left of fixation should show faster response times and less errors than scenes presented to the right of fixation. If however, the cortical locus of scene identification is in the left hemisphere then scenes presented to the right of fixation should show an advantage over scenes presented to the left of fixation.

Results

The results of Experiment 5 can be seen in Figures 16 and 17. A within participants t-test with Position (left visual field versus right visual field) as the independent variable was conducted in the first block data from Experiments 2 and 4. Separate t-tests were conducted

for the response time data and the error rate data. The t-test found no evidence that scene identification is subject to hemispheric specialization for response times, t (95) = .186, ns, SE = 12.23, or errors, t(95) = .976, ns, SE = .62. Power analysis determined that this experiment was capable of detecting a difference of 17 ms and 4.2% errors between the conditions.

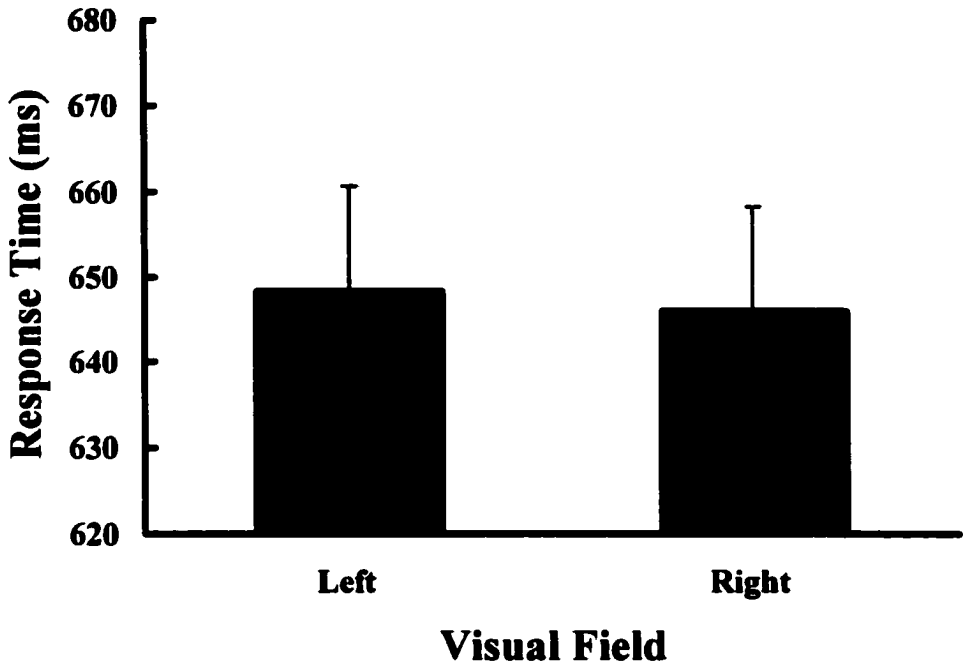

Figure 16. Mean response time to scenes presented to the left visual field and right visual field. Error bars represent the standard error of the mean.

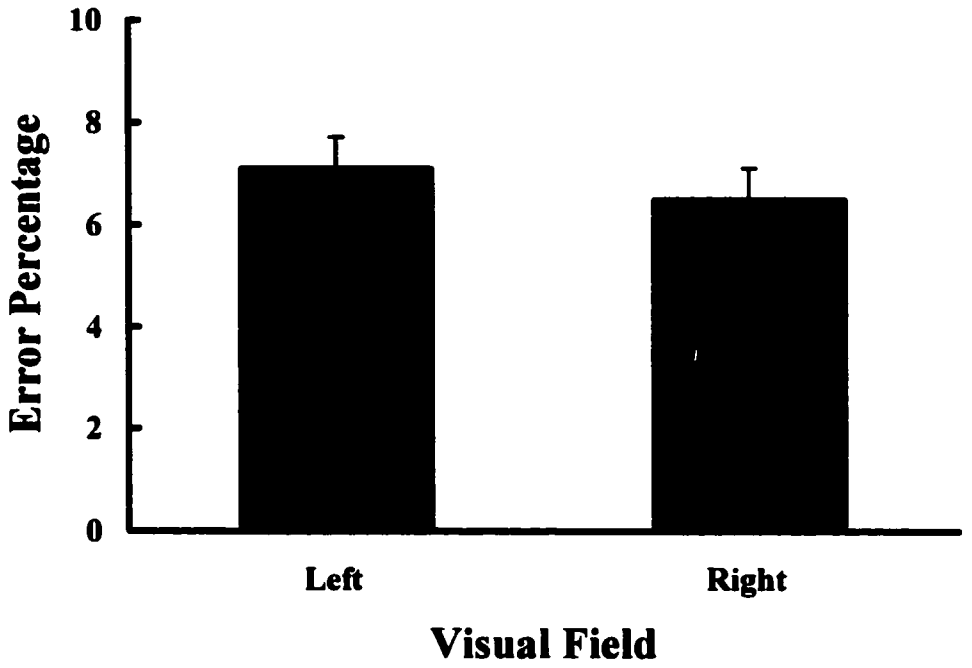

Figure 17. Mean error percentage to scenes presented to the left visual field and right visual field. Error bars represent the standard error of the mean.

Discussion

Experiment 5 found no evidence of hemispheric specialization for scene identification. Scenes that were presented to the left of fixation were identified just as quickly and accurately as scenes that were presented to the right of fixation.

One possible reason that null results were found in Experiment 5 may be because of the participants in the experiment. Laterality effects tend to be small and somewhat elusive (see for example, Kosslyn, 1994). Typically right handed males show the largest effects of lateralized presentation, and as a result some studies that report finding evidence of hemispheric specialization have limited their participants to only right handed males (see Springer & Deutsch, 1993). Therefore, the data from Experiment 5 were thus reanalyzed

with only the data from the 34 right handed males that participated in Experiments 2 and 4 included in the analysis.

Results

The results can be seen in Figures 18 and 19. Within-participants t-tests conducted on the response time and error rate data again revealed no evidence of specialization for response times, t (33) = .434, ns, SE = 19.67, or errors t(33) = .361, ns, SE = .95. Power analyses determined that this experiment was capable of detecting a difference of 21 ms and 4.2% errors between the two conditions.

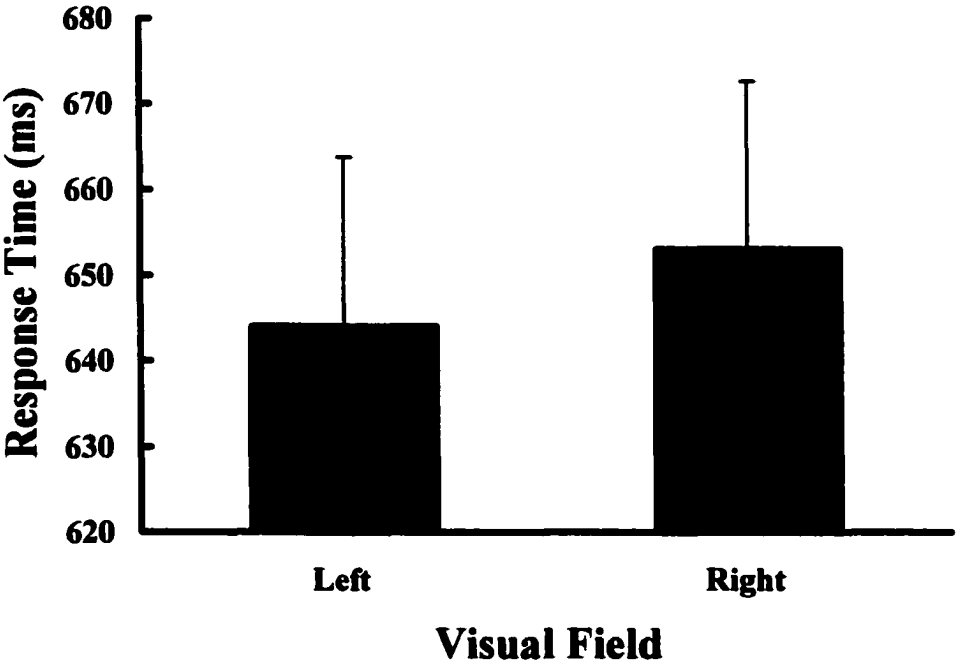

Figure 18. Mean response times for right handed males to scenes presented to the left visual field and right visual field. Error bars represent the standard error of the mean.

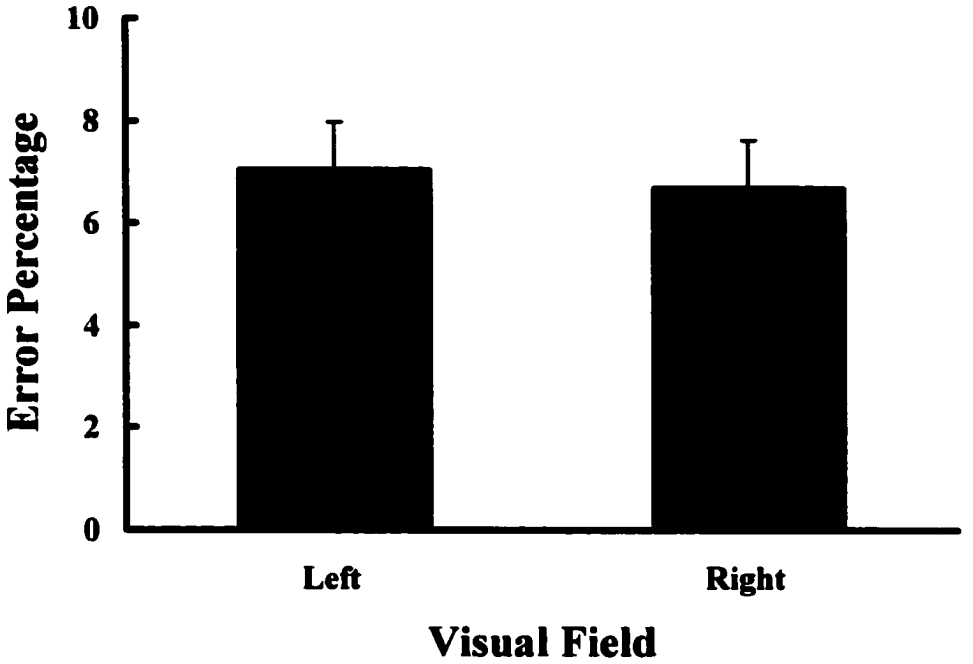

Figure 19. Mean error rates for right handed males to scenes presented to the left visual field and right visual field. Error bars represent the standard error of the mean.

Discussion

Neither analysis of the laterality data reveals an effect of visual field on the identification of scenes, even though a power analysis indicated the experiment was sufficiently powerful to detect a small difference in the hemispheres. As such, Experiment 5 found no evidence that hemispheric specialization occurs in scene identification. This is in contrast to other studies that have found a difference among the hemispheres in response to scenes (e.g., Epstein & Kanwisher, 1997).

There are several interpretations of the laterality results. First, it could be the case that there is no hemispheric advantage for scene identification, much in the same way that objects appear to be bilaterally represented (Biederman & E. E. Cooper, 1991b; Bryden &

Rainey, 1963; Levine & Banich, 1982; McKeever & Jackson, 1979). However, due to the null effect found in Experiment 5, converging evidence will need to come from other sources (such as studies of patients with brain damage or neuroimaging studies) in order to make this conclusion. Unfortunately, these studies have not yet been done.

Second, it is also possible that Experiment 5 was simply not sensitive enough to detect a hemispheric advantage. Although the power analysis indicated that a 21 ms difference could be detected, it is possible that a hemispheric advantage for scene identification could be smaller than this. However, as it stands, Experiment 5 found no evidence of a hemispheric advantage for scenes.

CHAPTER 7: GENERAL DISCUSSION

Summary of the Results

The results of the experiments reported here indicate that, unlike faces or objects, scene identification is sensitive to changes in the size and the position of the scene in the visual field. In Experiments 1 and 2, an identification advantage (priming) was only found when the scene was presented at an identical size and position on both blocks of trials. Changing the size of the scene was no better than presenting a completely new scene with the name, effectively eliminating visual priming. Similarly, changing the position of the scene reduced (but did not eliminate) priming as well. These results suggest that scene identification is not invariant to the same metric transformations as is object or face identification.

Likewise, the episodic recognition of scenes also appears to be sensitive to size and position transformations. In Experiments 3 and 4, participants found it more difficult to make an "old" response to scenes that were a different size or position than a scene presented in the first block of trials than to scenes that were an identical size or position as scenes in the first block. These results mirror the same metric sensitivity found for the episodic recognition of objects (Biederman & E. E. Cooper, 1992).

Finally, the results found no evidence that scene identification shows hemispheric laterality effects. Scenes that were presented to the left visual field (and thus initially processed by the right hemisphere) were identified just as quickly as scenes that were presented to the right visual field (and thus initially processed by the left hemisphere). This finding still occurred even when only right-handed males were included in the analysis.

Implications of the Results for Identification

The results of the experiments reported here indicate that the representation used for scene identification is different from the representation used for object identification and the representation used for face identification. Neither faces (Brooks, Rosielle, & E. E. Cooper, 2002) nor objects (Biederman & E. E. Cooper, 1991; Biederman & E. E. Cooper, 1992, Biederman, E. E. Cooper, & Hummel, 1992) show metric specificity across transformations in size and position. The results for faces and objects indicate that the same representation in memory becomes active when a face or object is viewed regardless of that face or object's size or position in the visual field. This metrically invariant method of coding is useful for faces and objects because size and position information is (in general) not useful for the identification of these classes of objects. For example, recognizing a watch does not depend on the size of the watch or where the watch is located in the visual field. The fact that the watch might be "big", or might be located in the upper left hand corner of the visual field is irrelevant information in recognizing the object as a watch.

Scenes, however, appear to be sensitive to changes in size or position, suggesting that this type of information is included in the representation used for identification. Two identical scenes, presented at different sizes or different positions in the visual field significantly reduced the amount of priming observed, sometimes eliminating visual priming altogether. The question, then, is what makes scenes different from faces and objects that they would be coded in such a way?

One obvious difference is that scenes, unlike faces and objects, can be navigated. Naturalistic scenes usually depict environments that a person could move through. In contrast, objects and faces are often stationary and this navigational component is almost

never present with individual objects and certainly never the case with faces. Further, people move through scenes, and as they move through a scene the visual input constantly changes, creating a very dynamic input. As such, one would need to constantly keep track of the changing position of many different elements (many of which are observed only in the periphery) in order to successfully negotiate the scene. Although the evidence regarding the extent to which metric information is actually used in scene navigation is mixed (see for example Presson & Hazelrigg, 1984), metric information does appear to be used for navigation under certain circumstances. Thus, one possible advantage of coding scenes metrically is that the same representation used for identification could also conceivably be used to navigate the scene. A single, metrically specific coding system could be used to both successfully navigate scenes and identify them. Studies that indicate the representation used for scene navigation is metrically specific (Christou & Bulthoff, 1999; Evans & Pezdek, 1980; Janzen, Schade, Katz, & Hermann, 2001; Pezdek & Evans, 1979; Wang, 1999) are consistent with this interpretation.

A second difference between scenes and other types of stimuli (faces and objects) is that scene identification involves computing the spatial relations between disparate, non-connected elements ("between-object relations", although it may not actually be discrete "objects" that serve as the primitives in scenes) whereas object and face identification involve computing spatial relations within an individual bounded entity ("within-object relations"). Studies have shown than between-object spatial relations tend to be more difficult to compute (Baylis & Driver, 1993), are computed by different brain regions (Humphreys & Riddoch, 1994, 1995) and tend to be more sensitive to metric changes (Saiki & Hummel, 1998) than within-object relations. The finding that scene identification is

sensitive to metric changes whereas object and face identification are not sensitive to metric changes might thus be explained by the differences in between versus within spatial relations.

How Well do Theories of Scene Identification Account for the Results?

Currently, researchers have offered two different hypotheses regarding the role of metric information in scene identification. Biederman (1988) proposed that scenes are identified using a metrically invariant memory representation. Alternatively, several authors (i.e., Diwadkar & McNamara, 1997; Shelton & McNamara, 1997) propose that scenes are identified with a metrically specific representation that includes both size and position information. Both of these hypotheses derive from extant theories of object identification and make very different predictions about the results of the current research. These hypotheses will be discussed in detail next.

Biederman (1998) proposed that scenes are identified via a set of volumetric primitive shapes ("geons," e.g., cones, cylinders, bricks) that combine together to form a metrically invariant representation of a scene. The results of the current research are clearly inconsistent with this theory. Although the current research does not rule out the possibility that Biederman's "geons" may be used as the primitives in scene identification (a notion consistent with the observation that a scene reduced entirely to geons can still be identified), the representation does not appear to be metrically invariant. Because Biederman proposes that the representation used to identify a scene does not include information about a scene's size or position in the visual field, his theory makes the prediction that no reduction in priming should be found when a scene changes size or position (because the same representation in memory would be activated regardless of size or position). However, the

results of the current research indicate that scenes that changed their size or position in the visual field did not produce any visual priming, indicating that two identical scenes of different sizes or two identical scenes presented at different places in the visual field active different representations in memory. Thus, Biederman's (1988) theory can not account for the result of these experiments.

In contrast, Diwadkar and McNamara (1997) (see also Shelton & McNamara, 1997) conclude that that metric information such as size and position are part of the representation used to identify scenes. The experiments reported here are consistent with this conclusion. Scenes that changed in size or changed their position in the visual field seemed to activate a different representation in memory for identification purposes. Further, the results here demonstrate that metric coding occurs in a task that explicitly tests scene identification (as opposed to the same/different discrimination task used in their experiments).

Although Diwadkar and McNamara do not specify the exact method of coding scenes metrically, nor do they specify a model of scene identification, they posit that models of object identification that use metric coding (e.g., Edelman & Bulthoff, 1992; Tarr, 1995; Ullman, 1989) can account for scene identification as well as object identification. Although the data for the models provide mixed results when applied to objects (see for example, Biederman & Gerhardstein, 1993; 1995), the results of the current research indicate that these theories may be applied to scene identification with some success.

However, not all models of identification that posit metric coding are supported by the results of the current experiments. It is possible to code the absolute position of the primitives in a scene (or object or face) and still achieve metric invariance in priming experiments such as the ones reported here. For example, Olhausen, Anderson and Van

Essen (1993) propose a model in which an attentional window determines the size and position coordinate frame by which the absolute positions of the primitives in the stimulus are coded. Because this coordinate frame can be different sizes and different positions (depending on where attention is allocated), a degree of metric invariance in achieved. For example, a large object in the upper right hand quadrant of the visual field would be coded in the same way as a small object in the lower left hand corner of the visual field, because at either size or position, an attentional window frames the object and the metric characteristics of the object are coded relative to his attentional frame. That is, if the attentional frame is wrapped around the object, the object might be coded as being "one unit tall" regardless of the object's absolute size.

Such a model is not supported by the data reported here. Like Biederman's theory, this model predicts that the changing the size or position of the scene should not affect the magnitude of the priming effect. Although the positions of the primitives of an object are coded metrically, changing the absolute size of the object (or scene) should have no effect on priming because the size of the attentional window is assumed to change as well. In the current research, however, changing the size and position dramatically reduced priming. Thus, it does not appear that an attentional window is used to achieve metric invariance, because such invariance is not achieved.

Implications of the Results for Models of Scene Processing

These experiments provide some of the first evidence of the nature of the memory representation used for identification and episodic recognition for scenes. These results also have the potential to inform theories of scene processing that make claims about the nature

of the memory representation formed when viewing a scene. In the following section I will discuss how the current set of results relates to these theories.

Rosielle, Crabb & Cooper's (2002) Categorical Coding

Rosielle, Crabb, and E. E. Cooper (2002) found evidence that categorical relations are being coded when viewing a scene; a finding that may seem to be at odds with the current research. Using a version of Rensink's (1997) flicker paradigm, Rosielle, Crabb, and E. E. Cooper found that changes to the position of an object in a scene that disrupted the object's categorical relationship to another object in the scene were detected more quickly than equal changes that only disrupted an object's metric relationship to another object. For example, changes that involved moving an object from "above" another object to the "side of" another object (a categorical change) were detected more quickly than changes that involved moving an object so that it remained "above" the other object (a metric change), even though the absolute distance moved by the displaced object was the same in both cases. These findings suggest that categorical relations such as "above" and "side of" are being coded in scene perception and that scene perception is relatively insensitive to metric changes (since these changes were more difficult to detect than non-metric (categorical) changes). If categorical relations dominate scene processing, and scene processing is insensitive to metric changes, then invariance to the size and position transformations used in the current research might be expected.

However, the results of Rosielle, Crabb, and E. E. Cooper (2002) do not rule out the possibility that metric coding also occurs in scene perception and the results of the current work do not rule out the possibility that categorical coding can occur as well. The current work does suggest, however, that these categorical relations might not be used for the

identification of scenes. In addition, attention is needed to detect change (Rensink, 2000; Rosielle, Crabb, & E. E. Cooper, 2002), and attention in scene perception is largely dictated by what is in foveal vision (see Henderson and Hollingsworth, 1999). Therefore, it could be the case that the attended region of the scene is given a higher degree of visual analysis, perhaps even creating a localized structural description (with categorical relations) of the attended (or foveal) region. This possibility is highly consistent with Rensink's (2000) model of change detection in scenes (discussed in more detail later), which proposes that rich and detailed local representations are temporarily formed at attended areas of a scene. These local representations allow change detection to occur.

However, it is unlikely that scene identification is entirely (or even mostly) based on information from the fovea. The fovea represents approximately 2% of the entire visual field, which presumably is too small of a region to encompass the information necessary to identify a scene. Further, although visual scanning can give one a more detailed representation of a larger area of the scene, numerous studies have found scene identification occurs prior to visual scanning taking place (Biederman, 1987b; Biederman , Mezzanotte, & Rabinowitz, 1982; Henderson & Hollingworth, 2000; Oliva & Schyns, 1997; Schyns & Oliva, 1994), and the exposure durations used in the current research were too short to allow visual scanning to take place. Further, certain types of scene categorizations can be made based on information displaced as much as 70.5° to the left or right of the fovea (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001) suggesting that it is possible for information in the periphery to play a role in scene identification. Thus, although there might be structural description-like coding occurring in the fovea (and the results of Rosielle, Crabb, & E. E. Cooper (2002) suggest that something like this does occur), much of the information

needed for scene identification would have to come from the periphery. If the representation created from the periphery is metric in nature, then this could account for the results of the current research, even though there might be metric invariance occurring in information from the fovea. Therefore, the current research and the results of Rosielle, Crabb, & E. E. Cooper (2002) may be consistent.

Rensink's (2000) Coherence Theory

Rensink (2000) outlined a model of how change can be detected in scenes. According to Rensink, early scene processing consists of the formation of "proto-objects", which are "relatively complex assemblies of fragments that correspond to localized structures in the world." (p. 22). Thus, proto-objects are not necessarily what are normally considered objects (such as cars and birds), but rather early precursors of detailed representations that are formed later in processing and include such properties as occlusion, curvature, slant, and shadowing. Rensink (2000) posited that these proto-objects are coded by their coordinates on the retina, are formed within a few hundred milliseconds, in parallel across the image, and do not require focused attention. These proto-objects provide the template from which focused attention can select regions of the scene to undergo further processing. Change detection occurs when a proto-object is selected (and processed) by attention both before and after the change takes place.

The results of the current experiments are consistent with Rensink's idea of a proto-object representation. The scenes in the current experiments were presented at exposure durations of a few hundred milliseconds and were found to be sensitive to metric changes for both identification and episodic recognition. If, as Resink proposes, a retinotopic map is

used to code the position of the primitives in scenes, then sensitivity to metric changes would be predicted. Thus, Coherence theory would predict the results of Experiments 1-4.

Coherence theory also proposed that these proto-object representations are volatile, meaning that in the absence of focused attention they quickly decay and are completely overwritten by new sets of visual input. According to Rensink, it is this focused attention, applied serially over different spatial areas of the scene, which can over time form permanent representations of the scene being viewed. However, without sustained focused attention, these representations are quickly and permanently lost.

However, the results of the current experiments suggest that, if they exist, these proto-object representations are not entirely volatile and are capable of influencing responses for some time into the future. Given that the participants in the current experiments did not have the luxury of sustained focused attention to various regions of the scene, Coherence theory would seem to predict that no visual priming should have occurred in Experiments 1 and 2. This was not the case; across both experiments an average of 62 ms of visual priming occurred after at least several minutes delay between prime and test. Although participants were certainly able to apply some focused attention to areas of the scene in the current research (especially in Experiment 1 when all scenes were presented in central vision), the attention was not sustained. Thus, although the current results are consistent with the notion that the early representation formed when viewing a scene is retinotopic in nature, they are inconsistent with the idea that these representations are completely temporary in the absence of sustained focused attention.

Two-Stage Processing Theories

Schyns and Oliva (1994; also Oliva & Schyns, 1997) proposed that scene identification occurs in two distinct stages; a "fast" stage in which the initial identification of the scene takes place, and a "slow" stage in which top-down processing (using the information from the fast stage) guides attention to form a detailed representation of the scene, including knowledge of the identity and positions of the objects present (see also Morrison, 2001). The information encoded during the fast stage is primarily low spatial frequency "blobs" oriented in a particular spatial organization. Similarly, Loftus, Nelson, & Kallman (1983) proposed that the initial fast stage of scene perception is based on "holistic" information, which (although poorly defined) seems to be the overall shape of the entire scene. This initial holistic representation activates one of a number of scene schemas, which is used to guide eye movements in order to process the detailed localized features of the scene.

The current results suggest that size and position information is included in these early representations (be they holistic or low spatial frequency blobs). This finding is consistent with theories of object processing that posit that early visual information is primarily metric in nature (e.g., Biederman, 1987; Hummel & Biederman, 1992; Rosielle & E. E. Cooper, 2000). For example, Rosielle and E. E. Cooper (2000) found that the orientation between the parts of an object was coded metrically, but only at short exposure durations. They argued that the metric coding was a product of representations formed early in the processing of visual information.

Implications for Laterality

Very little is currently known about the cortical locus of scene identification, although some clues are provided from studies that have used non-scene identification tasks.

Some studies point to the right hemisphere (e.g., Epstein & Kanwisher, 1998; Spiers et. al., 2001) whereas other studies (e.g., Bhattacharya & Petsche, 2002; Zaidel and Kasher, 1989) have found no hemispheric advantage for scene processing. Likewise, no hemispheric advantage was found in the current experiments. The scenes that were initially presented to the right of fixation (and thus the left hemisphere) were identified just as quickly as scenes that were presented to the left of fixation (and thus to the right hemisphere). This lack of a hemispheric advantage occurred even when only right handed males were included in the analysis. Finding no hemispheric advantage for scene identification is consistent with several studies done with objects that have also found non hemispheric advantage for object identification Biederman & E. E. Cooper, 1991b; Bryden & Rainey, 1963; Levine & Banich, 1982; McKeever & Jackson, 1979).

Implications for Priming

These experiments are the first to show that a class of stimuli (scenes) whose representation during identification shows a lack of invariance to metric changes. Although numerous experiments reported in the literature demonstrate metric invariance in the priming of object and face identification (e.g., Biederman & E. E. Cooper, 1991, 1992; Biederman & Gerhardstein, 1993, Fiser & Biederman, 1995; Fiser & Biederman, 2001; Hummel, 2001; Stankiewicz, Hummel & E. E. Cooper, 1998), no extant experiments have demonstrated lack of invariance for a particular stimulus class. Although Brooks, Rosielle & E. E. Cooper (2002) did find reduced priming for objects and faces after 180° picture-plane rotations, a 180° picture plane rotation changes non-metric information as well (see E. E. Cooper & Wojan, 2000, for a discussion).

The current results have implications for priming as a method to explore the visual representation underlying identification. These findings suggest that the invariance found in most other identification priming experiments is not an artifact of the priming paradigm itself. Given that all other repetition priming experiments have found no reduction in priming for size and position transformations, it might be the case that something about the priming paradigm itself, or the particular transformations used in the individual experiments, that led to the lack of invariance. For example, it could be that the visual system is simply insensitive to the particular size changes used in previous experiments. This does not seem to be the case, however, as the scenes used in the current experiments were presented at the same visual angle as those used in other experiments reporting invariance (e.g., Biederman & E. E. Cooper, 1992; Brooks, Rosielle, & E. E. Cooper, 2002). The current experiments thus help to validate priming as a means of testing for metric invariance.

Hummel (2001) proposed a model in which certain types of spatial relations (e.g, position and size) are coded in the absence of attention, leading to invariance to metric transformation that disrupt these relations. This model predicts that for visual priming, objects will show invariance to size and position transformations, regardless of whether they are attended or not. However, the current research demonstrates that this pattern does not hold for scenes, which showed reduced priming for both size and position. Thus, Hummel's (2001) model cannot account for scene identification.

Summary and Conclusions

The way in which people recognize the various elements of their environment is one of the fundamental questions in visual cognition. Although much is currently known about

the representation used to recognize objects and faces, very few studies have looked at the representation that allows scene recognition to occur.

Scenes appear to represent a different class of stimuli than faces or objects, at least for the purposes of identification. Past research has shown that both faces and objects are robust to metric transformations; however the representation used for scene identification appears to include both size and position information. No other experiments to date have found reduced priming after size and position changes for any type of stimulus, making the current set of results unique in the literature. Not everything in the visual environment is invariant to size and position transformations (for identification) as previously thought (see Hummel, 2001) and future models of identification will need to take this into consideration.

Further, scenes are not just large objects, as proposed by Biederman (1988). Unlike objects scenes appear to include information concerning the scene's size and position in the visual field. However, scenes are not entirely different from objects as both types of stimuli show the same pattern of metric sensitivity for episodic recognition. This finding suggests that although the representation used for identification is different for different classes of stimuli (e.g., faces, objects, and scenes) the representation used for episodic recognition may be similar. Thus, unlike for faces and objects, metric information plays a role in both the identification and episodic recognition of scenes.

REFERENCES

Allen, G. L., & Willenborg, L. J. (1998). The need for controlled processing in the visual acquisition of route knowledge. *Journal of Environmental Psychology, 18*, 419-427.

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented scenes. *Psychological Research, 43*, 277-292.

Archambault, A., O'Donnell, C., & Schynes, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science, 10*, 249-255.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*, 62-70.

Banich, M. T. (1997). *Neuropsycholgoy: The Neural Basis of Mental Function.* New York: Houghton Mifflin Company.

Barton, M. I., Goodglass, H., & Shai, A. (1965). Differential recognition of tachistoscopically presented English and Hebrew words in the left and right visual fields. *Perceptual and Motor Skills, 21*, 431-437.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects. Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 12-23.

Bhattacharya, J., & Petsche, H. (2002). Shadows of artistry: Cortical synchrony during the perception and imagery of visual art. *Cognitive Brain Research, 13*, 179-186.

Buswell, G. T. (1935). *How People Look at Pictures.* Chicago: University of Chicago Press.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177, 77-80.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual Organization.* (pp. 213-253). Hillsdale: Lawrence Erlbaum Associates.

Biederman, I. (1987a). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

Biederman, I. (1987b). Matching image edges to object memory. In *Proceedings of the First International Conference on Computer Vision,* (pp. 384-392). IEEE Computer Society, London, England, June, 1987.

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An Interdisciplinary Perspective* (pp. 370-487). Norwood: Ablex.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20,* 585-593.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 121-133.

Biederman, I., Cooper, E. E., & Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology, 46,* 191-214.

Biederman, I., Glass, A.L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97,* 22-27.

Biederman, I., & Gerhardstein, P. C., (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162-1182.

Biederman, I., & Gerhardstein, P. C., (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bulthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1506-1514.

Biederman, I., Rabinowitz, J. C., Glass, A.L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103*, 597-600.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 103*, 143-177.

Brooks, B. E., Rosielle, L., J., & Cooper, E. E. (2002). The priming of face recognition after metric transformations, *Perception*, *31*, 297-313.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception*, *23*, 803-822.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Bryden, M. P., & Rainey, C. A. (1963). Left-right differences in tachistoscopic presentation. *Journal of Experimental Psychology*, *66*, 568-571.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*, 1-47.

Christou, C. G., & Bulthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory and Cognition*, *27*, 996-1007.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology*, 81, 287-297.

Cohen, G., & Faulker, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, 4, 187-197.

Cooper, E. E., and Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic level object recognition. *Journal of Experimental Psychology: Learning Memory & Cognition, 26*, 470-488

Cooper, L. A., & Schacter, D. L. (1992). Dissociations between structural and episodic representations of visual objects. *Current Directions in Psychological Science, 1*, 141-146

Davidoff, J. (1982). Studies with non-verbal stimuli. In J. G. Beaumont (Ed.), *Divided visual field studies of cerebral organization* (pp. 29-55). New York: Academic Pres..

De Graef, P., De Troy, A., & D'Ydewalle. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychololgy, 46*, 489-508.

Delis, D. C., Robertson, L. C., & Efron, R. (1986). Hemispheric specialization of memory for visual hierarchical stimuli. *Neuropsychologia, 24*, 205-214.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damaged confined to the right hemisphere: An MRI and PET study and a review of the literature. *Neuropsychologia, 32*, 893-902.

Diamond, R., & Carey, S. (1986). Why faces are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Dickenson, S. J., Pentland, A. P., & Rosenfeld, A. (1993). 3-D shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*, 174-198.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302-307.

Edelman, S., & Bulthoff, H. H., (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32*, 2385-2400.

Ellis, H. D. (1983). The role of the right hemisphere in face perception, In A. W. Young (Ed.), *Functions of the right cerebral hemisphere* (pp. 33-64). New York: Academic Press.

Ellis, R., & Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioral study. In A. G. Cohen & J. R. Thomas (Eds.), *Artificial intelligence and its applications* (pp. 245-257). New York: Wiley.

Ellis, A. W., Flude, B. M, Young, A. & Burton, A. M. (1996). Two loci of repetition priming in the recognition of familiar faces. *Journal of Experimental Psychology: Learning Memory and Cognition, 22*, 295-308.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*, 598-601.

Evens, G. W., & Pezdek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 13-24.

Fiser, J., & Biederman, I. (1995). Size invariance in visual object priming of grey-scale images. *Perception, 24*, 741-748.

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation and hemisphere. *Vision Research, 41*, 221-234.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316-355.

Hellige, J. B. (1980). Effects of perceptual quality and visual field of probe stimulus presentation on memory search for letters. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 639-651.

Hellige, J. B., & Michimata, C. (1989). Categorization versus distance: Hemispheric differences for processing spatial information, *Memory and Cognition, 17,* 770-776.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243-271.

Henderson, J. M. (1992). Object identification in context. *Canadian Journal of Psychology, 46,* 319-341.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480-517.

Hummel, J. E., & Stankiewicz, B. J. (1996). Categorical relations in shape perception. *Spatial Vision, 10,* 201-236.

Humphreys, G. W., & Riddoch, M. J. (1994). Attention to within-object and between-object spatial representations: Multiple sites for visual selection. *Cognitive Neuropsychology, 11,* 207-241.

Humphreys, G. W., & Riddoch, M. J. (1995). Separate coding of space within and between perceptual objects: Evidence for unilateral visual neglect. *Cognitive Neuropsychology, 12,* 283-311.

Janzen, G., Schade, M., Katz, S., & Hermann, T. (2001). Strategies for detour finding in a virtual maze: The role of visual perspective. *Journal of Environmental Psychology, 21*, 149-163.

Kholmeyer, S. W. (1992). Picture perception lab: A program for picture experiments on the Macintosh II. *Behavior Research Methods, Instruments and Computers, 24*, 67-71.

Kosslyn, S. M., (1994). *Image and Brain.* Cambridge: MIT Press.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., & Koenig, O. (1992). Categorical versus coordinate spatial representations: Computational analysis and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 562-577.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli. (1989). Evidence for two types of spatial representations: Hemispheric specializations for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 723-735.

Levine, S. C., & Banich, M. T. (1982). Lateral asymmetries in the naming of words and corresponding line drawings. *Brain and Language, 17*, 34-43.

Levy, J., Trevarthen, C., & Sperry, R. W. (1972). Perception of bilateral chimeric figures following hemispheric deconnexion. *Brain, 95*, 61-78.

Loftus, G. R., & Loftus, E. F. (1988). *Essence of Statistics (2nd ed)*. New York: Random House.

Loftus, G. R., & Mackworth, N. H. (1974). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565-572.

Loftus, G. R., Nelson, W. W., & Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology, 35*, 187-198.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*, 43-75.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics, 2*, 572-552.

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive Neuropsychology: A Clinical Introduction*. New York: Academic Press.

McKeever, W. F., & Jackson, T. L. (1979). Cerebral dominance assessed by object- and color-naming latencies: Sex and familial sinistrality effects. Brain *and Language*, *7*, 175-190.

Metzger, R. L., & Antes, J. R. (1983). The nature of processing early in picture perception. *Psychological Research, 45*, 267-274.

Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. New York: Oxford University Press.

Mishkin, M., & Forgays, D. G., (1952). Word recognition as a function of retinal locus. *Journal of Experimental Psychology, 43*, 43-48.

Mishkin, M., Ungerleider, L. G., & Macko, A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417.

Morrison, D. J. (2001). Usage of spatial scales for the categorization of faces, objects and scenes. *Psychonomic Bulletin and Review, 8*, 454-469.

Nakatani, C., Pollatsek, A., & Johnson, S. H. (2002). Viewpoint-dependent recognition of scenes. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A*, 1115-139.

Olhausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13*, 4700-4719.

Oliva, A., & Schyns, P. G. (1997). Course blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72-107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176-210.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3*, 515-526.

Pezdek, K., & Evans, G. W. (1979). Visual and verbal memory for objects and their spatial locations. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 360-373.

Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 509-522.

Potter, M., & Levy, E. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

Presson, C. C., DeLange, N., & Hazelrigg, M. D. (1989). Orientation specificity in spatial memory: What makes a path different from a map of a path? *Journal of Experimental Psychology: Learning Memory and Cognition, 15*, 887-897.

Presson, C. C., & Hazelrigg, M. D. (1984). Building spatial representations through primary and secondary learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 716-722.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology, 46*, 342-376.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*, 17-42.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.

Reisberg, D. (1997). *Cognition: Exploring the Science of the Mind.* New York: Norton & Company.

Roediger, H. L., & Srinivas, K. (1993). Specificity of operations in perceptual priming. In P. Graf & M. Masson (Eds), *Implicit Memory: New Directions in Cognition, Development and Neuropsychology* (pp. 17-48). New York: Lawrence Erlbaum.

Rosielle, L. J., & Crabb, B. T. & Cooper, E. E. (2002). Attentional coding of categorical relations in scene perception: Evidence from the flick paradigm. *Psychonomic Bulletin & Review, 9(3)*, 319-326.

Rosielle, L. J., & Cooper E. E. (2001) The categorical perception of relative orientation in visual object recognition. *Memory and Cognition, 29*, 68-82.

Roskos-Ewoldsen, B, McNamara, T. P., Shelton, A. L., & Carr, W. (1998). Mental Representations of large and small spatial layouts are orientation dependent. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 215-226.

Saiki, J., & Hummel, J. E. (1998). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 227-251.

Schacter, D. L., & Badgaiyan, R. D. (2001). Neuroimaging of priming: New perspectives on implicit and explicit memory. *Current Directions in Psychological Science*, *10*, 1-4.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial scale dependent scene recognition. *Psychological Science*, *5*, 195-200.

Sergent, J. (1983). The role of the input in visual hemispheric asymmetries. *Psychological Bulletin*, *93*, 481-513.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain*, *115*, 15-36.

Sergent, J., & Signoret, J. L. (1992). Functional and anatomical decomposition of face processing: Evidence from prosopagnosia and PET studies of normal subjects. *Philosophical Transactions of the Royal Society of London*, *B335*, 55-62.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin and Review*, *4*, 102-106.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, *43*, 274-310.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156-163.

Srinivas, K., Breedin, S. D., Coslett, H. B., & Saffran, E. M. (1997). Intact perceptual priming in a patient with damage to the anterior inferior temporal lobes. *Journal of Cognitive Neuroscience*, *9*, 490-511.

Spiers, H. J., Burgess, N., Maguire, E. A., Baxendale, S. A., Hartley, T., Thompson, P. J., & O'Keefe, J. (2001). Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. *Brain, 124*, 2476-2489.

Springer, S., & Deutsch, G. (1993). *Left Brain, Right Brain* (4th ed.) New York: W. H. Freeman and Company.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 732-744.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society of London, 171 B*, 95-103.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A*, 225-245.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint-dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*, 55-82.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review, 2*, 339-363.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge from maps and navigation. *Cognitive Psychology, 14*, 560-589.

Thorpe, S. J., Gegenfurtner, K., R., Fabre-Thorpe, M., & Bulthoff, H. H., (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14*, 869-876.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 301-306.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193-254.

Wang, R. F. (1999). Representing the stable environment by egocentric updating and invariant representations. *Spatial Cognition and Computation, 1*, 431-445.

Winston, P. (1975). Learning strcutural descriptions from examples. In P. Winston (Ed.), *The Psychology of Computer Vision* (pp. 157-209). New York: McGraw-Hill.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The face that launched a thousand ships: Everyday difficulties and errors in recognizing people. *British Journal of Psychology*, 76, 495-523.

Yarbus, A. L. (1967). *Eye Movements and Vision.* New York: Plenum.

Zeki, S. (1999). The visual image in mind and brain. In. *The Scientific American Book of the Brain.* (pp. 17-28). New York: The Lyons Press.

Zaidel, D. W., & Kasher, A. (1989). Hemispheric memory for surrealistic versus realistic paintings. *Cortex, 25*, 617-641.

APPENDIX: LIST OF SCENE NAMES

The following is a list of target scene names used in Experiments 1 – 4.

Airport

Amusement park

Bar

Bathroom

Beach

Casino

Cave

Cemetery

Church

Circus

City

Classroom

Construction site

Desert

Farm

Forest

Freeway

Garden

Golf course

Kitchen

Living room

Mall

Market

Office

Park

Parking lot

Playground

Race track

Restaurant

Ruins

Street

Gym